# Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Stanford University*
Donald B. Lindsley, *Northwestern University*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Margaret K. Harlow, Managing Editor

## CONTENTS

*The Effects of an Interfering Task on the Learning of a Complex Motor Skill:* KATHERINE E. BAKER, RUTH C. WYLIE, AND R. M. GAGNÉ 1

*Food Deprivation and Discrimination Reversal Learning by Monkeys:* D. R. MEYER ..... 10

*Effect of Distribution of Practice on Rotary Pursuit "Hits":* R. B. AMMONS ..... 17

*Perceptual versus Analytical Responses to the Number Concept of a Weigl-Type Card Sorting Test:* D. A. GRANT ..... 23

*Transposition in Children as a Function of Age:* ELIZABETH ALBERTS AND D. EHRENFREUND 30

*An Experimental Investigation of Reactive Inhibition and Conditioned Inhibition:* K. C. MONTGOMERY 39

*Bidirectional Gradients in the Strength of a Generalized Voluntary Response to Stimuli on a Visual-Spatial Dimension:* J. S. BROWN, E. A. BILODEAU, AND M. R. BARON ..... 52

*The Influence of Two Variables upon the Establishment of a Secondary Reinforcer for Operant Responses:* P. J. BERSH ..... 62

*Problem Solving as a Function of the Situation:* M. R. MARKS ..... 74

American Psychological Association

Vol. 41, No. 1 January, 1951

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Avenue, N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for behavior theory. Normally, articles of a length exceeding 25 printed pages cannot be accepted. However, a single experimental study or an integrated series of studies accomplished simultaneously must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. All of the cost of the author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles must undergo the usual review by the Consulting Editors of the JOURNAL. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

The JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, including tables, footnotes, references, and titles for figures. Tables, figures, footnotes, references, and titles for figures must be placed on separate pages at the end of the manuscript.

---

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 41, NO. 1 JANUARY, 1951

# THE EFFECTS OF AN INTERFERING TASK ON THE LEARNING OF A COMPLEX MOTOR SKILL [1]

BY KATHERINE E. BAKER AND RUTH C. WYLIE

*Connecticut College*

AND

ROBERT M. GAGNÉ

*USAF Air Training Command,*
*Human Resources Research Center,*
*Lackland Air Force Base*

## INTRODUCTION

The purpose of the present experiment was to study the course of learning of a complex motor skill when an interfering task is introduced at different stages in the mastery of that skill. This was conceived of as a preliminary experiment which would provide descriptive data about performance in a situation requiring the mastery of a complex motor skill which is made up of two component tasks, a basic task and an additional interfering one. So far as we know, there are no such data in the literature. Such descriptive data would be of obvious practical value in planning motor skills training programs or in designing synthetic training devices. They are also of interest in that they present phenomena which must be accounted for by any theoretical explanation of motor learning.

In order to identify more definitely the kinds of learning situation for which our data have relevance, it is necessary to consider the various kinds of interference commonly arising in motor skills learning situations. Interference may occur in the form of masking or distorting of the signalling stimuli, it may consist in having to preform a basic task in a complex setting, or it may involve adding auxiliary tasks to the performance of the basic skill. It is this last type of interference with which the present experiment is concerned. The basic task to be learned was the manipulation of the S.A.M. Complex Coordination Test, in which the subject is required to align three sets of red and green lights by appropriate manipulations of a stick and rudder arrangement which resembles the controls of an airplane. After varying amounts

[1] This paper is a modified version of Report 316–1–9 under Contract N7onr-316, Task Order I, between Special Devices Center, Office of Naval Research, and Connecticut College.

of training on this task, the subjects were required to perform an additional task especially designed to interfere to a great extent with the basic skill. The measure of learning was the time taken to make successive groups of 26 settings on the Coordination Test.

From such an experimental setup it is possible to obtain descriptive data on several specific phenomena. First, we can see the course of learning of the basic skill before and after the introduction of the interfering task. In addition, by comparing groups just being introduced to the interfering task with those still practicing only the basic task, we can determine the relation between the amount of disruption of performance which occurs upon first introduction of an interfering task and the stage of learning at which interference is introduced. Finally, by comparing groups having had training on the component with those which have not, we may see whether positive transfer to the whole task results from practice on a component. In other words, the data should tell us the degree to which the facilitating effects of component practice may compensate for the disruption resulting from introducing the interference.

## Procedure

*Subjects.*—For this experiment, 132 young men volunteers were divided into six groups of 22, matched on the basis of the mean time taken for the first eight settings on the Coordination Test without the interfering task. These matching criterion scores were shown to correlate 0.49 ± .05 with the mean time for settings 27–52 on the Coordination Test, and 0.50 ± .10 with the mean time for Coordination Test settings 105–130.[2]

[2] The criterion score means ranged from 1.37 min. to 1.38 min., while criterion score *S.D.*'s ranged from 0.28 to 0.32. Records of a few subjects not used in the final matched groups were used in computing the correlations ($N = 95$ for the first correlation, 24 for the second).

*Apparatus.*—The basic task in the experiment was provided by the S.A.M. Complex Coordination Test (CM 701), which has been fully described by Melton (1). On the stimulus panel of the apparatus were three rows of red jewels and three adjacent rows of green jewels. On each trial one of the red jewels in each row was lighted, and the *S*'s task was to line up a movable green light with each of the red lights. Green lights in two of the rows were controlled by back-and-forth and left-right movements of a stick control, while the green lights in the third row were controlled by pressure of the right or left foot on the rudder bar. When all three red lights had been matched, a new setting of red lights automatically appeared, beginning a new trial. The model of the apparatus used in this experiment was the Standard Model E, with the substitution of less critical brush contacts which provided for the illumination of the green lights over 90 per cent of the distance between successive stationary contacts. This modification made the task easier, so that learning of the task was more nearly complete by the end of the 156 trials used in this experiment.

The main part of the apparatus for the interfering task was located on the floor at the left of *S*. It consisted of a covered wooden frame 24 in. in height, from which a gear shift handle protruded 4½ in. to a level slightly above that of *S*'s seat. The handle could be moved backward and forward in the slot through a total distance of 8 in. A red or a green jewel at the extreme left of the Coordination Test panel was lighted to provide a signalling stimulus for a response with the gear shift lever. Earphones were used to deliver instructions to *S* concerning what relationships between color of light and position of gear shift were to be the "correct" ones for the next few trials. The gear shift lever was connected with the wiring system of the Coordination Test in such a way that a new setting of red lights could not appear until the Coordination Test lights were aligned *and* the interference task gear shift lever was in the correct position.

A standard 1/100-min. timer was installed in the circuit in such a manner that it ran continuously from the time the first light setting appeared on the Coordination Test panel until *E* stopped it at the end of a group of 26 settings.

*Method.*—*S* was told, "We are going to ask you to learn to do this coordination task." *E* then sat down in the seat, showed *S* how to adjust the chair to suit body proportions, and gave one demonstration of how the lights were to be matched on the Coordination Test. Next, *S* was seated in the apparatus and allowed to make one setting of the lights. He was given all the help and instruction needed to make him understand the task.

Next he was given the following information about the interfering task. At various times during Coordination Test practice, instructions were to be given to him over the headphones, indicating the correct position of the gear shift lever. Any one of four instructions might be given: "red forward," "red back," "green forward," or "green back." This instruction was to be remembered until a new instruction was given. At any time during the performance of the Coordination Test, one of the lights at the extreme left of the Coordination Test panel might light up; and when that occurred, *S* was to be sure that the gear shift lever was in the position designated as correct by the most recent instruction. If the instruction said "red back" and a green light appeared, a forward position of the gear shift would be the correct one. Sometimes *S* might have to move the gear shift and sometimes not; but, in any case, a new setting on the Coordination Test panel would not appear until the lever was in the correct position and the Coordination Test lights were properly aligned.

*S* was then given the opportunity to make four practice settings of the gear shift, while *E* made corrections as necessary and made sure *S* fully understood the task.

Each *S* proceeded to practice the Coordination Test for 156 trials, which were divided into six sets of 26 settings each, with a 1-min. rest between each set. Each set of 26 settings, called learning stages 1 through 6, involved the same pattern of settings (Nos. 1 through 26), and provided one cumulated time score (in 0.01 min.) as a measure of the level of performance on the Coordination Test at that stage. No measurement has been made of performance on the interference task, but the degree of interference is reflected in an increase in time taken to make 26 settings on the Coordination Test in the presence of interference.

In addition to the time records for the six learning stages, a time score was taken for the first eight settings of Stage 1. These scores, obtained under non-interference conditions, later provided the basis for matching the experimental groups.

Although each of the six experimental groups went through all six learning stages, the interfering task was introduced at a different stage for each of the groups, as shown in Table I.

For the interfering task, one pattern of presentation of verbal instructions and light stimuli was made to correspond to each of the six learning stages, and all *S*s who were given interfering tasks at any given stage were given the pattern set up for that stage. Each of the six patterns involved: (a) interference task stimuli distributed over 26 settings (except for Stage 1 in which interference did not begin until the ninth setting); (b) four kinds of instructions, one of each of the four possible stimulus-response combinations; (c) four to six responses to the color mentioned in the instructions, and four to six responses to the color not mentioned explicitly by the instructions; (d) one to four presentations of stimuli to which the correct response was to make no movement; (e) one to four settings over which a single instruction must be remembered and acted upon one to four times; (f) presentation of interference stimuli one to three trials after the change to a new instruction.

TABLE I

SUMMARY OF EXPERIMENTAL CONDITIONS

| Group | Learning Stages without Interfering Tasks | Learning Stages with Interfering Tasks |
|---|---|---|
| I | None | 1–6 |
| II | 1 | 2–6 |
| III | 1–2 | 3–6 |
| IV | 1–3 | 4–6 |
| V | 1–4 | 5–6 |
| VI | 1–5 | 6 |

All groups, while in learning stages without interference, will be referred to as Non-Interference Groups; and all groups, while in learning stages with interference, will be referred to as Interference Groups. When a learning stage is accompanied by the interfering task it will be called an Interference Stage, whereas a learning stage not accompanied by the interfering task will be called a Non-Interference Stage.

## RESULTS

The means and standard deviations of the time scores for each group for the six learning stages are shown in Table II, and the means are presented graphically in Fig. 1. It may be seen that, at each stage of learning, all groups which have not as yet had the interfering task show essentially the same rate of learning, as measured by mean time scores.[3] On the first stage at which the interfering task is introduced, regardless of the stage of learning, the mean time score exhibits a sharp increase. This is followed on

[3] None of the critical ratios among the various possible pairings of these groups exceeds 1.00.

the next stage by a reduction in time score to about the level of the other Interference Groups.

Since the values for the Non-Interference Groups are so closely similar at each stage of learning, these values have been combined for use as measures of performance on the Coordination Test without interference. Thus, each Control Group is made up of all Non-Interference Groups at any particular stage of learning. The difference between the mean value from the Control Group and the value attained by an Interference Group on a given learning stage is employed as a measure of the amount of interference at that stage.

Analysis of the data shows that, at every stage, differences in time scores between Interference Groups and Non-Interference Groups are significant at the .05 level or better. Thus, we may say that, within the limits of the amount of training given in the present experiment, the detrimental effects of the interfering task are never completely overcome. It should be remembered that our interfering task was performed with the left hand, while the Coordination Test involved the use of the right hand. Thus, it is not inconceivable that with practice the two tasks might be performed simultaneously and not require any more time than for the settings of the Coordination Test alone. At any rate, if such effects do occur, the amount of practice in our experiment was insufficient to reveal them, since there was a significant degree of interference with performance on the Coordination Test, even after 130 practice trials.

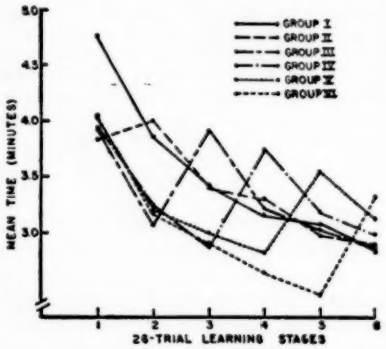


FIG. 1. Learning curves showing mean time in minutes taken to make successive groups of 26 settings on the S.A.M. Complex Coordination Test. Group I practiced an interfering task along with the Coordination Test during Stages 1–6; Group II practiced the Coordination Test on Stage 1 and then practiced on the Coordination Test with an interfering task during Stages 2–6; Groups III, IV, V, and VI practiced the Coordination Test during the first 2, 3, 4, and 5 stages, respectively, and then practiced the Coordination Test with an interfering task during the remaining stages.

It is evident from an examination of Fig. 1 that, for any given group, interference is at a maximum on the stage in which the interfering task is first introduced. The amount of this interference, at this initial stage, results in a performance score which is not only markedly less good than that of the Non-Interference (Control) Groups, but which is also inferior to the performance of any Interference Group which has already had one or more stages of practice on the combined tasks.[4]

The superiority of Interference Groups which are beyond the first Interference Stage is due to the fact that the amount of interference following the introduction of the interfering task diminishes rapidly with continued practice on the combined tasks. Thus, one stage of practice on the combined task is sufficient to

[4] Nine of the 15 differences between Interference Groups at the first Interference Stage and Interference Groups beyond the first Interference Stage, but at the same learning stage, are significant at the .05 level or better.

TABLE II

TIME IN MINUTES REQUIRED FOR SUCCESSIVE GROUPS OF 26 SETTINGS ON THE S.A.M. COMPLEX COORDINATION TEST WITH AND WITHOUT INTERFERING TASKS

| Group | Successive Groups of 26 Settings | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Group I | | | | | | |
| Mean | *4.768** | *3.850* | *3.425* | *3.186* | *3.085* | *2.848* |
| S. D. | *0.724* | *0.632* | *0.614* | *0.540* | *0.547* | *0.462* |
| Group II | | | | | | |
| Mean | 3.846 | *4.050* | *3.415* | *3.324* | *2.972* | *2.897* |
| S. D. | 0.688 | *0.691* | *0.510* | *0.647* | *0.534* | *0.511* |
| Group III | | | | | | |
| Mean | 3.935 | 3.072 | *3.927* | *3.243* | *3.047* | *2.866* |
| S. D. | 0.812 | 0.523 | *0.767* | *0.558* | *0.462* | *0.608* |
| Group IV | | | | | | |
| Mean | 3.958 | 3.252 | 2.870 | *3.749* | *3.201* | *2.966* |
| S. D. | 0.845 | 0.819 | 0.525 | *1.039* | *0.583* | *0.603* |
| Group V | | | | | | |
| Mean | 4.010 | 3.232 | 3.005 | 2.838 | *3.531* | *3.165* |
| S. D. | 0.659 | 0.579 | 0.486 | 0.491 | *0.629* | *0.471* |
| Group VI | | | | | | |
| Mean | 4.063 | 3.172 | 2.898 | 2.654 | 2.478 | *3.331* |
| S. D. | 0.886 | 0.593 | 0.479 | 0.439 | 0.459 | *0.694* |
| Control Groups † | | | | | | |
| Mean | 3.963 | 3.182 | 2.923 | 2.745 | 2.478 | |
| S. D. | 0.796 | 0.643 | 0.525 | 0.451 | 0.458 | |

* Interference trials shown in italics.

† Control groups consist of all *S*s who had not as yet had inteference, i.e., Groups II–VI at Stage 1, Groups III–VI at Stage 2, Groups IV–VI at Stage 3, etc.

bring the performance of any Interference Group up to the level of performance of other Interference Groups which have had larger proportions of their previous practice on the combined tasks.[5] It is clear, therefore, that, although introducing the interfering task makes the performance of any group suffer considerably, this disadvantage, compared to other Interference Groups, is in all cases overcome by a small amount of additional training on the combined tasks.

The data show that both the *absolute* amount and the *relative* amount of increase in time score, which accompany the first introduction to the interfering task, become progressively greater with increasing numbers of trials of Coordination Test practice. Fig. 2A shows the relationship between the percentage increase in time for 26 settings and the amount of previous practice on the Coordination Test without interference. The values along the ordinate represent percentage increases in time scores for each group relative to Control Group scores. Inspection of this figure shows that when the interfering task is introduced immediately, time scores are about 20 per cent higher than those obtained for groups which undertook the Coordination Test without interference. By the fifth learning stage, interference amounts to a 42 per cent increase in time over what would be expected at this stage

[5] None of the possible comparisons between Interference Groups beyond their first Interference Stage is significant at the .05 level or better.

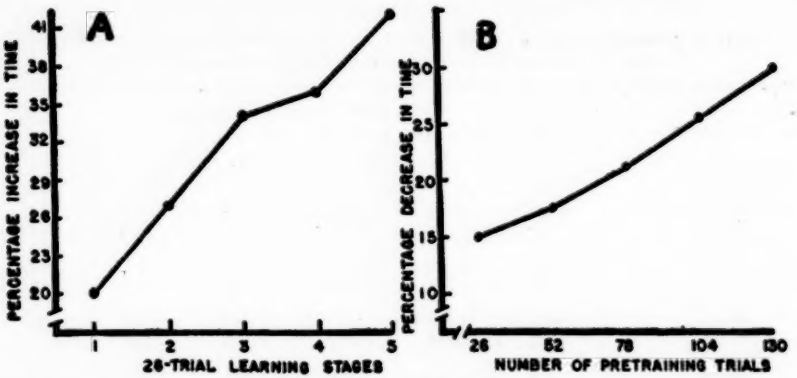


FIG. 2. A. Interference as indicated by percentage increase in mean time for 26 settings on the Complex Coordination Test under newly introduced interference conditions as a function of the learning stage at which an interfering task was first introduced. Each datum point represents the percentage increase in time for one of the experimental groups over what would be expected at the particular stage had the interfering task not been introduced.

B. Transfer effects from varying amounts of component task practice to performance in the first learning stage at which the interfering task was introduced. Transfer is measured by the percentage decrease in time scores for the Coordination Test with interference tasks, from what would be expected had no practice on the Coordination Test alone been given. Thus, each datum point represents a comparison of the performance of a different experimental group with the performance of Group I on Stage 1.

had no interference been introduced.[6] This increasing function shows that a second interfering task brings about a relatively greater disruption of performance with increasing numbers of training trials on the component task alone.

Considering the experimental problem as a study in transfer of training, we may determine how much transfer to the total task (Coordination Test plus interfering task) is brought about by increasing amounts of practice on the basic Coordination Test alone. From Fig. 1 it may be noted that as the amount of training on the basic task is increased, there is also a regular decrease in the time score on the total task at the first Interference Stage. The amount of this reduction in time score following increased amounts of practice on the Coordination Test may be expressed as a percentage decrease from the score for Group I on Stage 1, which is the level of performance attained on the combined tasks when no previous training (except instructions) has been given. The precentage reduction from this level attained by each group can be taken to indicate the amount of positive transfer to the combined tasks which has been brought about by training on the Coordination Test alone. Fig. 2B is a graphic representation of this relationship. It is evident that 26 trials on the Coordination Test alone bring about a 15 per cent improvement in time score on the total task. Each subsequent 26-trial practice unit is accompanied by an

[6] No statistical tests for the significance of the differences between these successive percentages have been made. But the fact that they are a continually increasing function and double in value from the first to the fifth learning stage, suggests that this trend was not due to chance alone.

additional 4 per cent reduction in time.[7]

Considering the information contained in both Figs. 2A and 2B, we may say that, although a second task results in greater interference when introduced in the later stages of learning, the amount of positive transfer from practice without interference likewise increases. Thus, an increase in interference is more than compensated for by the concomitant increase in transfer effected by the additional component practice.

## Discussion

These results indicate that an interfering task of the type used in this experiment produces a sharp increase in the time necessary to perform the Coordination Test, and that, within the number of trials of practice used here, these interferences are never completely overcome as compared with performance without interference. It is also shown that Interference Group curves tend to be superimposed after the first Interference Stage has been passed. Although a second task becomes increasingly interfering when its introduction is postponed, increasing positive transfer effects are shown by the fact that the scores on the first Interference Stages become progressively lower with increasing numbers of practice trials on the Coordination Test alone.

The preliminary nature of the present study does not permit us to identify the factors which are responsible for the phenomena described. However, considering first the increased amounts of interference found with greater amounts of Coordination Test practice, we may reasonably speculate that part of the explanation lies in the high degree of similarity between both stimuli and responses of our interfering and coordination tasks.

If it is assumed that continued practice on the Coordination Test has the effect of progressively strengthening particular stick and rudder response tendencies to red and green lights of the kind used in this task, such practice may be expected to result in increasing strengths of generalized response tendencies to similar stimuli. The lights used in the interfering task may be considered highly similar to those of the Coordination Test in both color and location. Thus, it may be supposed that, at the time when the interfering task is introduced, the *S* who has had a great deal of practice on the Coordination Test will have a greater tndency to move the stick and rudder when presented with the interfering task stimuli. *S*s with less Coordination Test practice would have less of this tendency, and disruption of performance would be smaller. If this speculation is correct, the degree to which disruption of performance occurs would depend upon similarity relationships between the stimuli of the two tasks. In addition, since the gear shift movements closely resemble the stick movements, there may also be some tendency to make movements with the gear shift which would be the correct stick movements; such a tendency would also increase with greater amount of practice on the Coordination Test. The degree to which such interferences occur would depend upon the degree of similarity between responses of the tasks. A test of the correctness of this reasoning lies in further experiments em-

[7] At the beginning of learning, 26 trials on the component task yield a significant increment in transfer; later in learning, 78 or more additional trials on the component task are required to yield a statistically significant increment in transfer.

ploying interfering task stimuli and responses which are decreasingly similar to the basic task stimuli and responses.

Let us now consider a possible theoretical explanation for the increasing positive transfer resulting from increasing amounts of component practice. This finding is entirely consistent with the notion that increasing positive transfer will result from continued practice which seemed to decrease progressively the internal generalization within the Coordination Test. Such a reduction in internal generalization means that discrimination of the Coordination Test stimuli will be more easily accomplished and responses will be better defined. The rate of increase in discriminability of stimuli and decrease in variability of responses may be great enough so that interference effects contribute less and less to the net result in performance of the Coordination Test. Again, the suggested hypothesis must await further evidence for its verification.

A further word of comment should be made concerning the interpretation of the findings of this experiment. The present results are based on a time measure of learning and do not necessarily indicate trends that might appear if errors had been used as the index of learning. Also, gross units of practice were used in measuring performance on the Coordination Test, so that the course of learning within any given stage is unknown. Finally, since only 156 practice trials were given, it is impossible to tell how much training must occur before transfer and interference effects are no longer increasing functions.


## Summary

1. An experiment is reported which deals with the effects of introducing an interfering task at various stages during the learning of a motor skill. The motor skill used here was the S.A.M. Complex Coordination Test. The interfering task consisted of appropriate forward and backward movements of a gear shift arrangement in response to the appearance of a red or green light. The experiment involved six matched groups of 22 college men. The learning was divided into six stages of 26 trials each. One group had the interfering task from Stage 1 onward, another from Stage 2 onward, etc.

2. The interfering task produced an increase in the time necessary to make settings on the Complex Coordination Test. At each stage of learning there was an initial sharp increase in time when the interfering task was being introduced for the first time at this stage. Although the time score was reduced as practice was continued on the combined tasks, it never reached the level of the curve for subjects who had not been given the interfering task.

3. Time scores for the Interference Groups at the various learning stages were compared. With the exception of the group receiving interference after only 26 settings, each group, while in its first Interference Stage, gave a time score which significantly exceeded the mean times of other groups which had practiced with interference over one or more stages prior to that stage. After the first Interference Stage was passed, all Interference Groups continued to improve at essentially the same rate

regardless of the proportion of interference and non-interference training.

4. Increases in time scores upon the first introduction of interference were absolutely and proportionately greater with increased amounts of practice on the Coordination Test alone.

5. The amount of positive transfer from non-interference training also increased. Therefore, despite increasing interference, the absolute level of performance became increasingly better with greater numbers of training trials on the Coordination Test alone.

(Manuscript received
September 24, 1949)

REFERENCE

1. MELTON, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program Research Reports, Report No. 4. Washington, D. C.: Government Printing Office, 1947.

# FOOD DEPRIVATION AND DISCRIMINATION REVERSAL LEARNING BY MONKEYS[1]

BY DONALD R. MEYER[2]

*University of Wisconsin*

This experiment is concerned with the relationships between intraproblem and interproblem reversed discrimination learning by monkeys, and with the influence of variations in time of food deprivation upon the acquisition of the discrimination reversal learning set (3). During the presentation of a typical problem *Ss* are first trained to select one of two markedly dissimilar objects. Subsequently, the reward values of these objects are reversed. Both the course of learning on the reversed discrimination trials and its modification with practice on a series of equivalent problems have been demonstrated to have the orderliness and predictability ordinarily found only in the most circumscribed behavioral situations.

Since the formation of the discrimination reversal learning set alters performance in a problem itself thought to involve complex interactions, we might expect it to be a complicated process; but the functions that have been obtained are relatively simple. At present, however, the relationships between the magnitude of the interproblem increment in efficiency and the course of intraproblem learning are incompletely understood. This paper will develop a coherent, quantitative theory of reversal learning which will relate these two aspects within a common framework.

In almost all investigations of primate learning, preferred foods are used as rewards for test performance. Experimental sessions at the Wisconsin Primate Laboratory are typically scheduled at from 13 to 23 hrs. after the animals are fed with a distinctly different, but nutritionally adequate, maintenance diet. The consequences of changes in the drive that has its strength defined in terms of these variations have received little systematic study. Whether or not the food that is obtained in the test situation is relevant for this drive cannot be readily ascertained, but it is possible that facilitation can be accomplished either by relevant or irrelevant drives. The latter belief—in varying terminology—is widespread. This experiment is designed to check an empirically developed working assumption to the contrary, i.e., that large variations in time of maintenance food deprivation are without effect upon the levels of performance maintained by the special test foods.

## PROCEDURE

*Subjects.*—From a group of 12 rhesus monkeys, all previously tamed and adapted to the test environment, the nine *Ss* required by the statistical design were chosen (Numbers 137, 140, 142, 143, 146, 147, 149, 150, and 151). The other three animals were kept in reserve to insure a full table of data, but their use was not required.

In an unpublished food deprivation experiment, the animals had been trained in discrimination learning. A series of 168 pair of multidimensional stimulus-objects (e.g., rubber heel vs. matchbox) had been presented, each for a

[1] This experiment is part of a dissertation conducted under the direction of Professor Harry F. Harlow and submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology at the University of Wisconsin in 1950.

[2] Now at Ohio State University.

total of six trials. *Ss* had learned, with increasing success, to choose the consistently rewarded object of each pair. The monkeys received the maintenance diet at 1:30 or 2:00 P.M., and testing followed at 3:00 or 3:30 P.M., 8:00 or 8:30 A.M., and 12:00 or 12:30 P.M., respectively. No significant differences in performance were found.

*Apparatus.*—The Wisconsin General Test Apparatus (3, Fig. 1) was used throughout. It consists of a table placed in front of a barred restraining cage. Mounted on the table is a superstructure holding two movable screens, one of which is opaque and located just in front of the cage. The other, a one-way vision screen, is in front of *E*, who sits at the table opposite the cage. A movable tray containing two small food-wells can be advanced toward and withdrawn from the cage. The wells are covered during test trials by the stimulus-objects.

The restraining cage floor was marked off to facilitate an activity sample, and a latency apparatus was devised which consisted of a Standard Electric Chronometer in series with a switch on the forward screen and two switches on the test tray. The latter switches were closed by the weight of the objects, and the former by a contact which closed when the screen was raised.

*Experimental design.*—The experimental series consisted of 108 reversed discrimination learning problems, each of which utilized a different pair of multidimensional stimulus-objects. The data were collected by three highly trained *Es* who followed a mutually developed and standardized testing technique.

The three levels of food deprivation studied were 1 hr., 23 hrs. and 47 hrs. None of these periods was accompanied by overt signs of disorganization. It was assumed that the 6, 8, or 10 trials given on each problem before the introduction of the reversal would assure near-maximal development of the antagonistic discrimination regardless of the prevailing deprivation conditions.

Predetermined left-right sequences established the positions of the stimulus-objects on the tray. Each of 12 basic sequences was represented in each of nine lists. These basic sequences were constructed for six prereversal and eight reversal trials, and to them zero, two, and four prereversal trials were added. Each of the resulting balanced lists was made up of four sequences of six, four of eight, and four of ten prereversal trials. Each sequence was used for one reversal problem, and the 12 sequences of a given list appeared in random order.

The statistical design, which is diagrammed in Fig. 1, was based on a 9 × 9 greco-latin sqaure constructed by randomly selecting and permuting two squares of the orthogonal set published by

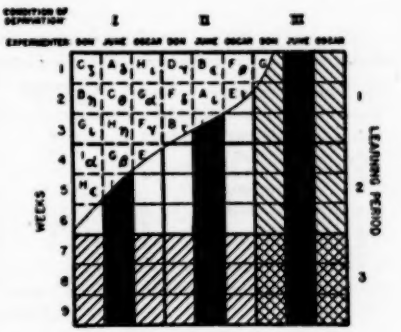


FIG. 1. Diagram of the statistical design

Fisher (1). *Ss* were assigned to the latin square, problem lists to the greek. Interproblem changes with learning appeared on the nine rows, each row representing one week's experimentation. The three deprivation periods and three *Es* were assigned to the nine columns in an arrangement which preserved the deprivation × experimenter interaction. In the figure a single cell represents data from 12 reversal problems, and monkey, list, week, *E*, and deprivation conditions are determined by the letters in the cell and the labels of row and column. The contribution to total variance of all the cited variables was removed from error with a relatively minor loss in degrees of freedom. The general properties of this and other latin square designs were recently summarized by Grant (2).

*Living and feeding arrangements.*—The monkeys were housed in pairs selected to assure clear-cut but moderate dominance. Animals of the 1- and 23-hr. groups were fed daily at 12:00 noon. Two food pans, each of which contained 160 gm. of Purina Laboratory Chow and seven 10-gm. pieces of carrot, were placed and anchored at mutally isolated positions in the cage. Characteristically the dominant monkey went immediately to one pan, the submissive monkey to the other. The animals were permitted to eat for 20 min., and at 12:25 P.M. each was given 60 gm. of orange.

All monkeys were tested three times a week on alternate days. On Saturdays the pans were left for 30 min. as an additional precaution against long-range nutritional changes. The monkeys were invariably finished at the end of this period. The food deprivation period for the 47-hr. group of the week began at this time. Two days later, following testing, another 30-min. feeding period was given.

The testing period for all groups was relatively compact, and diurnal differences unrelated to

TABLE I

SUMMARIES OF ANALYSES OF VARIANCE

| Source of Variation | d. f. | Prereversal Errors *F* | Reversal Errors *F* | Total Activity *F* | Forward Activity *F* |
|---|---|---|---|---|---|
| 1. Food deprivation | 2 | 1.76 | 1.92 | <1 | <1 |
| 2. Experimenter | 2 | 2.20 | 3.74* | 18.13*** | 31.54*** |
| 3. Deprivation × Experimenter | 4 | <1 | <1 | <1 | <1 |
| 4. Weeks (Learning) | 8 | 1.34 | 21.86*** | 1.30 | 1.26 |
| 5. Monkeys | 8 | 11.75*** | 33.30*** | 15.43*** | 13.89*** |
| 6. Lists | 8 | <1 | 1.11 | 1.58 | 1.95 |
| Error Mean Squares | 48 | 20.75 | 15.73 | 314.33 | 180.49 |

Confidence level: * .05; *** .001

feeding were minimized. The 25-hr. group was tested at 11:00 A.M., the 47-hr. group at 11:30 A.M. and the 1-hr. group at 1:30 P.M. In the latter group cheek pouches were generally clear of food by about 10 min. before the beginning of experimentation.

*Testing method.*—Each monkey was brought to the apparatus room from the living quarters in a small transport cage. Upon arrival the animal was given five pieces of test food, peanuts or raisins, and the testing began. Both screens were lowered, the forward screen was raised, and after a fixed delay of 2 sec. a 30-sec. time sample was taken of entrances into each of the cage quadrants. The forward opaque screen was then lowered and the first trial of the first reversal problem arranged. The one-way vision screen was lowered to a position which just permitted manipulation of the tray. After the tray was advanced, the forward opaque screen was raised. The monkey reached through the bars of the restraining cage to choose one of the multidimensional stimulus-objects covering the food wells on the tray. After six, eight, or ten prereversal trials, the food was placed under the previously incorrect object. This and the subsequent seven trials were discrimination reversal trials, the ones in which we were primarily interested. On each day five time samples were taken and four problems completed. The monkey was then given five additional pieces of test food and returned to the living quarters.

Separate scoring was made of prereversal and reversal errors, mean prereversal and reversal latency, and forward and total restraining cage activity. Experimenter intercorrelations on the latter observations, determined at the beginning of the experiment, were high (.90+). A trial was defined as the displacement of either of the stimulus-objects far enough to uncover any portion of the underlying food-well.

## RESULTS

The summaries of the analyses of variance for error and activity data are presented in Table I, and mean scores for three of the measures are graphed in Fig. 2. A high frequency of delays

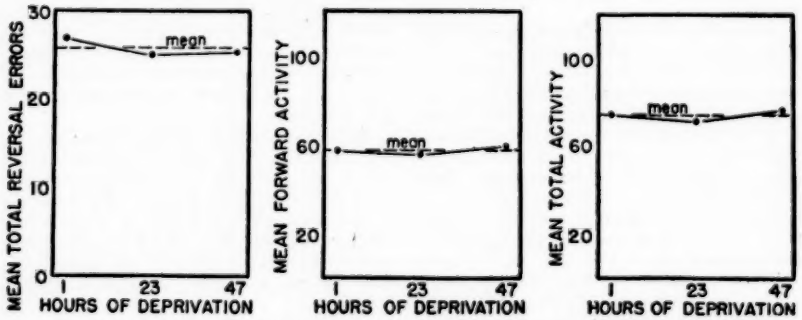


FIG. 2. Mean error and activity scores as a function of time of maintenance food deprivation

beyond an arbitrary limit of 5 sec. rendered indeterminate the original plan for scoring latency data. An analysis of the frequency of such delays was attempted, but these data were so unreliable that differences as great as those between animals could not be demonstrated.

For error and activity scores the lists main effect and the deprivation × experimenter interaction were uniformly absent. Monkey differences, as expected, were large on all measures. Experimenter variance was significant on all measures except prereversal errors, but the differences in the reversal errors were not marked.

No differences associated with variations in time of maintenance food deprivation appeared in either the activity or error scores. This cannot be interpreted as a lack of sensitivity, for the curves of Fig. 2 are effectively flat. The uniform failure of differences suggests that, over the range studied, maintenance food deprivation is an unimportant variable.

The most detailed analysis was made of variations with learning in the reversal error measure, and these data gave results of considerable theoretical interest. Scores were combined into three-week groups, Learning Periods 1, 2, and 3, to give the mean intraproblem error data presented in Table II. The course of intraproblem learning at the beginning of the experiment was estimated from the 36 problems of the sessions for the first day.

TABLE II

PER CENT ERRORS, TRIALS 1–8

| Trial | Stages of Interproblem Learning | | | |
|---|---|---|---|---|
| | First Day | Learning Period | | |
| | | 1 | 2 | 3 |
| 1 | 92 | 90 | 91 | 91 |
| 2 | 72 | 46 | 24 | 19 |
| 3 | 58 | 36 | 22 | 16 |
| 4 | 42 | 30 | 18 | 13 |
| 5 | 33 | 19 | 15 | 9 |
| 6 | 33 | 20 | 13 | 7 |
| 7 | 14 | 15 | 10 | 10 |
| 8 | 22 | 16 | 6 | 8 |

Except for the First Day data, which are continuous throughout, there is a break in the intraproblem curves at the second trial point. This discontinuity grows progressively more abrupt with continued problem practice. For these latter curves the segment from the first to the second trial is described as the *fast rate* portion, and the remainder as the *slow rate* portion.

The ready application of similar asymptotic decay functions to the First Day data and to the slow rate portions of the other curves suggested that *all these data were described by a single general equation.* The simplicity of this description was interpreted as futher evidence for the intraproblem discontinuity, and the determination of the equation by empirical methods was undertaken.

It had been shown (3, Fig. 8) that the plotting of per cent errors on the second trial as a function of *linear* amounts of problem practice resulted in a relationship closely approximating the asymptotic exponential form. Since these scores represent initial points of unplotted slow rate functions, this curve intimated that optimal overlaps of the obtained intraproblem error curves would be obtained if it were assumed that equal amounts of problem practice translated the segments equal distances along a trial-unit abscissa. Inspection of the data confirmed the feasibility of this approach. The data for Learning Periods 2 and 3 were superimposed upon data for Learning Period 1 by translations of 2.6 and 5.2

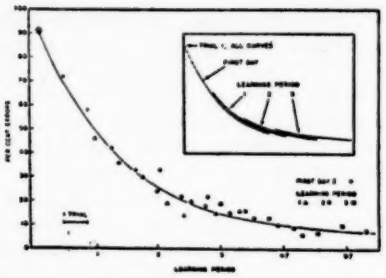

FIG. 3. A graph of the general reversal equation and the overlapping intraproblem functions from which it was derived

trial units respectively. The function for the First Day was located, appropriately, 1.3 units to the left of that for Learning Period 1.

To the data replotted in this way, and with the Trial 1 value estimated at 91, least squares fits were made using varying asymptotes. The finally accepted equation, in semilogarithmic form, is

$$\log (y - 5.6) = 2.0603 - .1261x.$$

A graph of this equation and the points from which it was derived are given in Fig. 3. Because of the linear relationships between abscissal translation and amount of problem practice, a transformation of this equation involving $x$ alone can easily be made so that the value of $y$ can be computed from values of $x$ expressed in learning period units.

Although individual differences between monkeys were great, the basic similarities between the forms of individual curves and the summary curve can be appreciated through examination of Fig. 4. The data from all 108 problems were used in estimating these functions.

## DISCUSSION

The present experiment concurs with the cited preliminary study in substantiating the assumption of investigators in the Wisconsin Laboratory that performance for special test foods is unrelated to time of maintenance food deprivation. Over a very extended range of values this variable is without effect in the discrimination and discrimination reversal situations. It does not modify the orientation of the monkey in the test situation, the over-all activity score, or the error scores. The drive, then, that is so defined has few, if any, of the characteristics usually ascribed to drives.

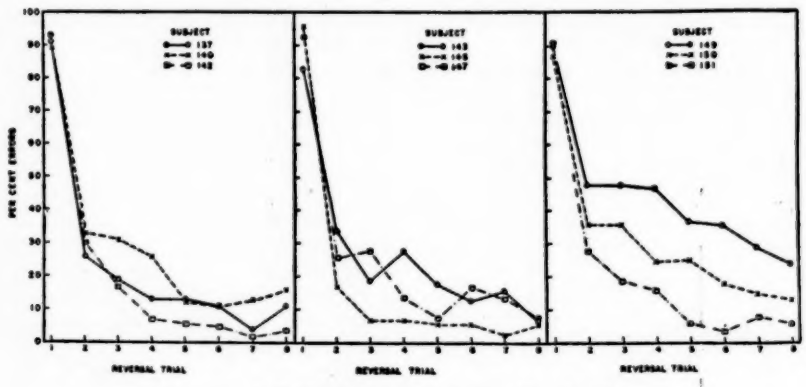

FIG. 4. Individual differences in intraproblem functions

As learning progresses on a series of discrimination reversal problems, the curve for intertrial performance on a particular problem divides into two distinct parts. Improvement takes place at a rapid rate from Trials 1 to 2, and at a slow rate on all subsequent trials. The intraproblem curve, then, acquires a progressively greater discontinuity.

The slow rate portions of all intraproblem reversal learning curves can be described by a single general equation, with the rate proportional to the discrepancy between obtained and limiting per cent error performance. The demonstrated independence of the form of the intraproblem slow rate curve and magnitude of the interproblem improvement is not readily discerned through casual inspection, but it is clear that the steep functions obtained in the early problems and the flat functions that appear in the late problems are the rapid and the asymptotic portions of a single nonlinear curve.

As practice on successive reversal problems proceeds, the slow rate functions are initiated progressively closer to the asymptote. In this way the interproblem increment determines how much of the range of improvement will be traversed at the inefficient rate but does not modify the slow rate function itself. Interproblem learning takes place, rather, by a substitution of rapid improvement for slow.

Although the formation of the reversal learning set results in a discontinuity in the intraproblem learning curves, the change in the function of the initial, experimenter-induced negative trial is itself both orderly and continuous. This interproblem improvement has been demonstrated to have essentially the same form for both monkey and preschool child subjects (3).

With the aid of the obtained quantitative interrelationships, a description of the entire course of discrimination reversal learning—both interproblem and intraproblem—can be derived from a very limited amount of appropriately chosen data. The interproblem equation, for example, can be estimated from as few as two intraproblem functions and expressed in terms of units defined by the amount of problem practice separating the two intraproblem curves. Such relationships obviously permit a large number of very specific predictions.

From a reliable reversal error function chosen anywhere in the training series an equation can be derived which will fit, with some portion of its course, any reversal error function. It is similarly possible to estimate, from a given score, the performance on the succeeding and preceding trial (provided the latter is not Trial 1). For this prediction the trial or problem position of the offered datum is not required. It should be noted finally that the resulting improvement on successive problems with the development of the interproblem increment can be expressed in terms of *a saving of trials* as the rapid drop encroaches upon the slow rate function. Measured in this way the improvement is a linear function of the number of problems encountered.

## Summary and Conclusions

This investigation analyzed the relationships between intraproblem and interproblem discrimination reversal learning by monkeys, and studied the effects on performance for special test foods, of variations in the length of maintenance food deprivation.

Performance for special food rewards was not influenced by duration

of maintenance food deprivation. The periods studied were 1, 23, and 47 hrs.

Interproblem improvement in reversal learning was found to result from the growing differential action of the initial reversal trial.

The form of the function describing intraproblem learning was independent of the development of the reversal learning set. Although a discontinuity developed in these curves as problem practice progressed, the continuous portions obtained were at all times fitted by a single asymptotic decay function.

Interproblem improvement, measured in terms of savings of trials, was a linear function of the amount of problem practice.

(Manuscript received
December 17, 1949)

## References

1. Fisher, R. A. *The design of experiments.* New York: Hafner, 1947.
2. Grant, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
3. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

# EFFECT OF DISTRIBUTION OF PRACTICE ON ROTARY PURSUIT "HITS" [1]

BY ROBERT B. AMMONS

*University of Louisville*

In the past, the typical method of measuring rotary pursuit performance has been to hook some sort of timer or counter into the stylus-target circuit and to record for given time intervals the total time of completion of the circuit. This record does not give information as to the temporal patterning of the contacts or "hits" during practice. Wing and Ammons (**7**) developed a method of recording this temporal pattern of circuit completions on a moving tape, from which information could be obtained for the individual *S* as to number of hits, total duration of hits per trial, mean duration of hits, and mean duration of misses.

The author has applied this method of analysis to the tape-recorded performances of *S*s practicing continuously, with a single rest (**2**). It was found that during the initial period of continuous rotary pursuit, the number of hits slowly increased, variability in number of hits remained fairly constant, mean duration of hits and its variability increased slowly, and mean duration of misses and its variability decreased rapidly. As it has been shown (**1, 6**) that markedly better total rotary pursuit performance results when rest pauses are introduced between trials, it is to be expected that hit performances would also differ between groups with massed and distributed practice. Perhaps a study of these differences in hit patterns would throw some light on the gross phenomena of distribution of practice.

[1] This study was made possible by grants from the Carnegie Foundation Research Fund and the Tulane University Council on Research. The author wishes to thank Mr. Cliff Wing, who aided in the collection of the data, and Mr. Joseph R. Murphy, who made most of the computations.

## Problem

It was the purpose of this study to determine the effects of distribution of practice early in rotary pursuit on number of hits, variability in number of hits, mean duration of hits, variability in mean duration of hits, mean duration of misses, and variability in mean duration of misses.

## Procedure

*Apparatus.*—The apparatus consisted of four 60 r.p.m. pursuit rotors with ¾-in. brass targets whose centers were set 3¼ in. from the centers of 10½-in. black vinylite disks. Styluses had freely swinging arms, rounded silver tips, and were so constructed that the weight on the tip was .6 oz., while the total weight was 3.6 oz.

Six V DC stylus-target circuits operated relays which in turn controlled the current to two banks of four .001-min. Standard Electric timers and the magnetic ink-writing pens of a polygraph. Trial timing was done manually by means of a master switch which switched the recording from one bank of four timers to the other and actuated the polygraph's magnetic ink-writing, time-marking pen.

Recording tape moved past the polygraph pens at a constant speed of 4.5 in. per 20-sec. trial. One pen recorded a continuous line for each of the rotors, showing stylus-target circuit completions as excursions from a base line.

*Subjects.*—*S*s were two groups of ten undergraduate women college students, naive to the rotary pursuit task. They volunteered from psychology classes, and were assigned, as available, to practice groups of four. There were no known variables which systematically biased the sample in either condition. These *S*s were part of larger groups run in an experiment reported elsewhere (**3**).

*Method.*—Ss were told beforehand of the general nature of the task and that the study was of the way skills were learned. They were instructed to try to keep the stylus tip in contact with the target as much as possible during practice. Following was to be done with an easy, relaxed, swinging movement of the body, and no extra pressure was to be put on the stylus tip in any way. The stylus handle was to be held lightly in the preferred hand with a tennis-type grip, which was demonstrated. Performance deviating from these instructions during practice was corrected verbally.

Ss practiced in groups of four, standing approximately 3 ft. apart, back to back, in a V with *E* at the point. Distributed practice groups were "sandwiched," one group practicing during another group's rest periods. Encouragement was given periodically during practice, and at the end, a general statement was made about the level of performance reached by each *S*.

One group of ten Ss practiced 36 22-sec. trials without rest, while the other group of ten rested 5 min. between each pair of 22-sec. trials. The performance during the first 2 sec. of each trial was not recorded, so scores are those made during 36 20-sec. trials for each group. The initial 2-sec. period was introduced to minimize starting difficulties for the distributed practice group, which was given a 20-sec. warning to take up rotor position and a 2-sec. warning that pursuit was about to be resumed. During their rest periods, Ss talked, read, and drank soft drinks. The 5-min. rest interval was chosen on the basis of other experiments (**1, 3**) as leading to a nearly maximum rate of acquisition and level of performance. Thus the two conditions provided for high degrees of massing and distribution of practice.

## Results

Records were read individually with a 6-power wide-field binocular microscope against a 1/100-in. scale. These figures were later converted to seconds of total possible stylus-target circuit completion time. A portion of a record is reproduced by Wing and Ammons (**7**), who describe the mechanics of record reading in some detail. Twelve to 15 hours were required to read and evaluate each record, which process included recording the duration of each hit and each miss and calculating the values

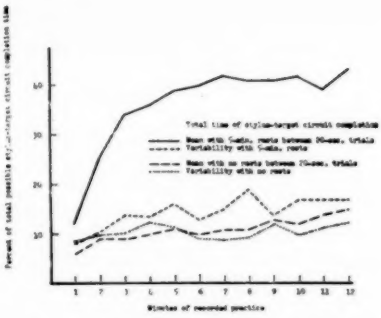

FIG. 1. Mean total time and variability of total time of stylus-target circuit completion per 20-sec. trial as affected by distribution of practice

of the various indices used in group computations.

Fig. 1 gives the mean total time and variability of total time of stylus-target circuit completion per 20-sec. trial. *Total time* was computed for each *S* for each 20-sec. trial as the per cent of time she kept the stylus and target in electrical contact. These values were averaged for the group by sets of three 20-sec. trials to obtain the index of *mean total time*. This index is similar to that of mean clock time per trial as used in other studies. The index of *variability of total time* was computed by (a) obtaining the variance of total times within 20-sec. trials, (b) averaging these variances three trials at a time, (c) taking the square root of this figure.

It can be seen from Fig. 1 that the mean total time of the continuous practice group shows a slow, almost linear increase, while the mean total time of the distributed practice group shows a more rapid, negatively accelerated increase to a much higher level. This finding agrees with results in other studies (**1, 3**) using mean clock time per trial. Variability of total time increases slowly for both

groups, with that of the distributed group being somewhat greater. The variability for the massed group is probably limited by the smaller mean total time.

Fig. 2 presents an analysis of the number of hits per 20-sec. trial. Hits lying totally within a given trial period were counted as *whole hits*, while those overlapping the start or stop of a 20-sec. trial period were counted as *partial hits*. *Number of hits* was calculated for each *S* for each 20-sec. period, a whole hit counting as 1 and a partial hit as ½. Numbers of hits were averaged by 20-sec. trials for the whole group, then the 20-sec. means averaged by groups of three to give greater stability to the index. *Variability of number of hits* was calculated by (a) obtaining the variance of number of hits for the group for each 20-sec. trial, (b) averaging these variances by groups of three, (c) taking the square root of this figure.

Inspection of Fig. 2 reveals that the mean number of hits by the continuous practice group rises slowly throughout practice, while that by the distributed practice group rises sharply to a plateau of about 32 hits per 20-sec. trial within 4 min. of practice. Both groups start practice

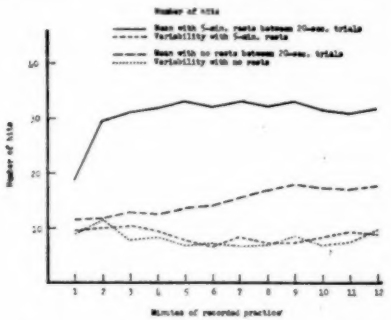

FIG. 2. Mean number of hits and variability of number of hits per 20-sec. trial as affected by distribution of practice

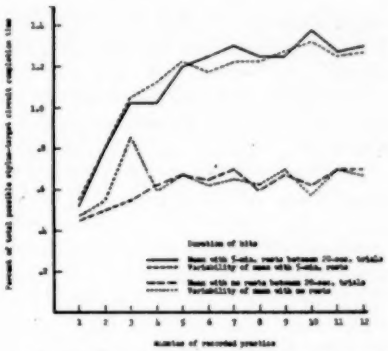

FIG. 3. Mean duration of hits and variability of mean duration of hits per 20-sec. trial as affected by distribution of practice

at considerably above the zero level. It should be noted that number of hits is limited, since better performance involves hits of greater duration as well as more hits. Maximum performance would consist of one hit of 20-sec. duration per 20-sec. trial. Thus it is to be expected that number of hits ultimately would have decreased if more practice had been given. Variability of number of hits, as shown in Fig. 2, is fairly constant and nearly equal for the two groups throughout practice.

Fig. 3 presents the analysis of hit duration for the two groups. For each *S* an *individual mean duration of hits* was calculated for each 20-sec. trial by averaging the durations of all hits during a given trial and counting a whole hit as 1 and a partial hit as ½ in obtaining an *N*. *Mean duration of hits* was the mean for the whole group of these individual mean durations of hits by 20-sec. trials. These means were averaged three at a time to give a more stable figure. This method of calculation gives equal weight to the individual mean duration rather than to the individual hit.

*Variability of mean duration of hits* was computed by (a) obtaining the

*mean duration of whole hits* for each *S* for each 20-sec. trial by averaging the durations of whole hits only, (b) calculating the variance for the group for each 20-sec. trial of these mean durations of whole hits, (c) averaging these variances by groups of three to give more stable figures, (d) taking the square root of the three-trial average. It will be noted that all hits were used in the calculation of mean duration of hits, while it was felt that only whole hits could be used meaningfully in the computation of variability of mean duration of hits.

In Fig. 3 we see that mean duration of hits increases during both massed and distributed practice. Duration is considerably greater for the distributed practice group but starts well above zero for both groups. Both curves are apparently negatively accelerated. Variability of mean duration of hits in both cases is nearly equal to mean duration. This probably does not mean much except that variability is limited by the mean values, and the distribution is highly skewed. The high value at minute 3 for the massed group was due to one very extreme score which did not affect the mean duration as much as the variability.

The last analysis made was of duration of misses. The results are shown in Fig. 4. *Mean duration of misses* and *variability of mean duration of misses* were computed in the same way as the corresponding duration of hit indices, except that durations of individual misses were used as raw scores. Inspection of Fig. 4 shows that mean duration of misses decreased markedly for both groups during practice, and at all times was less for the distributed practice group. This group shows a sharper decrease in mean miss duration and reaches a stable point after about 5 min. of practice, beyond which there is no further improvement.

The massed practice group improves more slowly, and miss duration is still decreasing at the end of practice. The variability of mean duration of misses curve closely follows the mean duration curve for each group and probably has little quantitative significance beyond its indication that the distributions are markedly skewed.

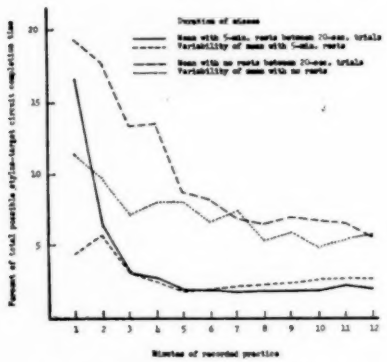

FIG. 4. Mean duration of misses and variability of mean duration of misses per 20-sec. trial as affected by distribution of practice

## DISCUSSION

The present group differences can be attributed to massing and distribution of practice only if the groups were initially comparable in performance. The skewness of the distributions of the values of almost every index precludes the use of ordinary small sample statistics. Inspection of the performance of the two groups during the first 20-sec. trial showed that the massed group was slightly superior with respect to mean duration of hits, mean duration of misses, and number of hits, while the distributed group was slightly superior with respect to mean total stylus-target contact time. It was apparent

that there were no significant differences between the groups initially.

Special note should be taken in Fig. 4 of the failure of the mean duration of misses for the distributed group to improve or decrease after approximately 5 min. of practice. A miss roughly represents the duration of a combination visual and kinesthetic reaction time to the start of a movement of the stylus tip off the target plus the duration of the corrective movement. It would seem that this adjustment cannot be made, on the average, in less than about .4 sec. It is interesting to note that Searle and Taylor (5) found that it took from .41 sec. to .50 sec. to make a similar correction during linear tracking. As the present task can be considered as rotary tracking, a direct comparison can be made. The agreement between the two studies is striking.

Some light is thrown on Kimble's two-factor theory of inhibition (4). Kimble believes that "if the rest periods in the work situation are too short to allow for the dissipation of the majority of the reactive inhibition (as in the case of massed practice learning), $I_r$ accumulates and there develops an increasingly strong tendency (need) for the organism to cease work or rest." Once a critical level of reactive inhibition (need to rest) is built up, resting will occur until the amount of reactive inhibition drops below the critical level. Thus an equilibrium will be set up.

The present data support this position in a general way. *S*, with continuous practice, leaves the target sooner on the average (starts to rest?), and stays off longer (rests longer?). To be testable rigorously with respect to rotary pursuit, Kimble's theory will, of course, need operational definitions of its constructs in terms of that task.

Certain speculations can be made about the nature of the rotary pursuit response and its acquisition. We know that performance improves as *S* learns to make longer and longer unbroken sweeps and to correct more quickly for deviations from these movements which would take him off the target. In addition, it is known from direct observation that misses are recorded by this particular apparatus when the stylus tip is not held close to one spot on the target even though it stays on the target. At the start of practice, *S* is able to maintain his stylus tip steadily in one position on the target for about one-half a revolution. As he improves in distributed practice, he averages over one time around between successive breakings of contact. He also reports initially that the target follows an irregular path.

It is proposed that *S* learns to make longer and longer sweeping, ballistic-type movements of both eye and hand and that he also learns to correct "errors" in a shorter time, both because he is sensitive to more cues than initially and because he has learned more efficient corrective movements. *S* starts out with some ability to make smooth ballistic-type circular movements. Improvement takes place as he is able to make longer and longer movements without direct "voluntary" control. By ballistic-type, non-voluntarily-controlled movements is meant sweeping movements where change of direction takes place in less than normal reaction time and is consequently not under direct cortical control.

## Summary and Conclusions

Twenty undergraduate college women practiced rotary pursuit for 36 22-sec. trials, ten working continuously and ten resting 5 min. between

trials. A polygraph record was made for each *S* showing when the stylus was held steadily on the target, thus completing an electrical circuit. Such a completion was called a hit, and the non-completion time between two successive hits was called a miss. A series of hit and miss indices was defined and differences between the groups described.

The groups were initially comparable in performance. It was found that: (a) *number of hits* is greater during distributed practice; (b) *number of hits* rises quickly to a rather stable maximum level during distributed practice but increases slowly throughout continuous practice; (c) *mean duration of hits* is greater during distributed practice; (d) *mean duration of hits* shows a negatively accelerated increase during practice for both groups; (e) *mean duration of misses* is less during distributed practice, decreasing quickly to a stable minimum of about .4 sec.; (f) *mean duration of misses* decreases slowly throughout massed practice.

It is concluded that the hit indices described differentiate clearly between performance during massed and that during distributed practice. Variability of these indices is in most cases limited by the mean values and so cannot be meaningfully compared. The learning of steady circular pursuit movements may well involve the building up of manual and ocular movement patterns of a ballistic type, the components of which are not under direct cortical control.

(Manuscript received
December 20, 1949)

## References

1. Ammons, R. B. Acquisition of motor skill: III. Effects of initially distributed practice on rotary pursuit performance. *J. exp. Psychol.*, 1950, **40**, 777–787.
2. Ammons, R. B. An analysis of "hits" in continuous rotary pursuit with a single rest. *J. gen. Psychol.* (In press.)
3. Ammons, R. B. Relationship of motivation and method of testing to distribution of practice phenomena in rotary pursuit. *J. exp. Psychol.* (In press.)
4. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
5. Searle, L. V., and Taylor, F. V. Studies of tracking behavior. I. Rate and time characteristics of simple movements. *J. exp. Psychol.*, 1948, **38**, 615–631.
6. Spence, K. W., Buxton, C. E., and Melton, A. W. *The effect of massing and distribution of practice on rotary pursuit test scores.* Civil Aeronautics Administration, Division of Research, Report No. 44, 1945. Washington, D. C.: U. S. Dept. Commerce.
7. Wing, C. W., and Ammons, R. B. Measuring rotary pursuit performance by kymographic recording. *Motor Skills Res. Exch.*, 1949, **1**, 9–12.

# PERCEPTUAL VERSUS ANALYTICAL RESPONSES TO THE NUMBER CONCEPT OF A WEIGL-TYPE CARD SORTING TEST [1]

BY DAVID A. GRANT

*University of Wisconsin*

The problem of the present study was to compare sorting for number when the stimulus configuration was constant for each number with sorting for number when the stimulus configuration was *not* constant for each number.

The Wisconsin Card Sorting Test (WCST) was used in the experiment (1, 2, 3). The test materials consist of stimulus and response cards. Each response card was printed with one, two, three, or four identical figures—triangles, stars, crosses, or circles—in one of four colors—red, green, yellow, or blue. The 64 response cards consisted of all possible combinations of these. The four stimulus cards—one red triangle, two green stars, three yellow crosses, and four blue circles—were placed in the upper set of four double compartments on a sorting tray, and *S* was instructed to sort the response cards under the stimulus cards and to try to get as many "right" as possible. He was told that the experimenter would tell him on each sorted card whether he was "right" or "wrong."

The cards could easily be sorted for the color, number, or form of the figures. *Color* was first called "right" by *E* until *S* had sorted ten successive cards correctly. *E* then changed to *form*, calling all *form* responses "right" and all *color* or other responses "wrong." Thus the only cue to the shift was the change in *E*'s system of calling "right" and "wrong." After *S* gave ten successive correct *form* responses, *E* shifted again, this time to *number*, and, after ten successive correct responses here, went back to *color*, *form*, and, finally, to *number* in the same way, making six successive stages in all.

In a previous experiment (3) it was discovered that college students found it easier to sort for *number* than for *form* or *color* on the WCST. The superiority of number sorting was found to occur because the *S*s acquired the number sorting response more readily and perseverated longer in making number responses. Although it was not supposed that the particular *color*, *form*, and *number* concepts[2] of the test were representative of *color*, *form*, and *number* concepts in general, it was, nevertheless, expected that the results would be more in line with Heidbreder's hypothesis that *number*, requiring some analysis by *S*, should be more difficult than *color* and *form*, which are more directly perceived (4, 5). Her findings in modified memory or "concept formation" experiments with similar concepts (6, 7, 8, 9, 10, 11) revealed that number concepts were delayed and difficult. In card sorting behavior Heidbreder found that sorting for *number* typi-

[1] This research was supported in part by the Research Committee of the Graduate School of the University of Wisconsin with funds voted by the State Legislature for 1948-49. The writer is indebted to the Computing Service of the University of Wisconsin for some of the statistical analysis.

[2] The term concept will be used throughout this paper in a purely descriptive sense, referring only to the objective regularities in the stimulus and response cards. This usage is similar to that of Heidbreder (6).

cally occurred later than sorting for *objects*, *spatial arrangements*, *forms*, or *linear qualities* (**12**, pp. 14–18) unless situational support or special sets favored numerical sorting (**12**, pp. 19–23).

Analysis of the standard WCST suggested that conventional and configurational aspects supported the number concept. For example, the stimulus cards were arranged across the board, one, two, three, four, in a conventional left to right arrangement. Furthermore, the *one* figure was always centered, the *two* figures were always along the principal diagonal of the card, the *three* figures always formed an equilateral triangle arrangement, and the *four* figures always formed a square arrangement. *S* could, therefore, sort according to configuration rather than according to number, and fragmentary verbal reports seemed to bear this out.

The present experiment was designed to test the hypothesis that these conventional and configurational supports would materially increase the ease of number sorting in an abstraction situation such as that provided by the WCST.

## Procedure

*Materials.*—Two sets of standard stimulus and response cards and two sets of special response cards were used. On the special cards the figures were printed in random spatial arrangements. The standard cards, because of the systematic placement of the figures, will be referred to as systematic (*S*) response cards, and the special cards, because of the unsystematic placement of the figures, will be referred to as unsystematic (*U*) response cards.

*Experimental design.*—Three factors were varied in the present experiment, arrangement of stimulus cards, use of *S* or *U* response cards during the regular six stages of the WCST, and transfer to *S* or *U* response cards on an additional seventh stage of the test, in which *S* was also to sort for number. The stimulus cards were arranged two ways: the standard or systematic (*S*) way and an unsystematic (*U*) way—three yellow crosses, one red triangle, four blue circles, and two green stars, respectively, from left to right across the sorting tray. The three factors, stimulus card arrangement, response cards, and response cards on the transfer stage, were combined in a trifactorial design as shown in Fig. 1. Each of the eight groups thus required may be identified by a three-letter combination referring to the factors in order: stimulus, response, transfer. Thus the *SUU* group would have *S*-stimulus card arrangement, *U*-response cards on stages 1 to 6, and would continue with *U*-response cards on the seventh or transfer stage. The *USU* group would have *U*-stimulus card arrangement, *S*-response cards for stages 1 to 6, and would be handed a fresh deck of *U*-response cards for the seventh stage, being instructed at that point to continue doing what they were doing.

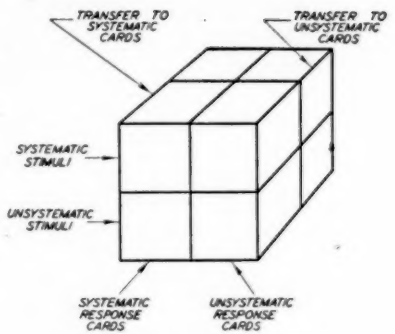


Fig. 1. Design of the experiment

*Subjects.*—The *S*s were 200 men and women students who volunteered from classes in elementary psychology at Northwestern University and the University of Wisconsin. Twenty-five were assigned to each group according to the order of their appearance. In addition, 33 *S*s were run and eliminated because they required more than 64 cards to complete a single stage. All *S*s were run singly in a small, sound-treated room or a quiet office. There were no significant differences between performances of the Northwestern and Wisconsin *S*s.

## Results

*Scoring.*—Each *S*'s performance at each stage of the experiment was recorded in terms of three scores: (1) The number of *perseverative errors* at each stage except the first and seventh

—this was the number[3] of incorrect responses which would have been correct on the immediately preceding stage; (2) the number of *non-perseverative* errors—these responses would also have been wrong on the preceding stage of the test; and (3) the *total correct* responses—this was the number of correct responses made by *S* at each stage[4] before he began his criterion sequence of ten successive correct responses. The larger the number of correct responses, the lower was *S*'s efficiency in shifting from the old response to the new correct response.

*Data on first six stages.*—The average number of perseverative errors, non-perseverative errors, and total correct responses for each group are presented separately for each stage of the experiment, in Table I. The table shows some general trends. The number of perseverative errors tends to drop off gradually from the second stage to the sixth stage. The number of non-perseverative errors drops off more rapidly from the first stage to the seventh. The total correct increases from stage 1 to stage 2, dips down at stage 3, then rises again for stages 4 and 5, and finally drops down on stages 6 and 7.

The effects of the three factors of the experiment (or their interactions) may be examined by dividing the eight groups of Table I appropriately. For example, the effect of the differences between *S*- and *U*-response cards can be found by averaging together the totals of the *SSS*, *SSU*, *USS*, and *USU* groups to obtain the data for the *S*-response cards, and averaging the totals of the *SUS*, *SUU*, *UUS*, and *UUU* groups to obtain the

TABLE I

PERSEVERATIVE ERROR, NON-PERSEVERATIVE ERROR, AND TOTAL CORRECT SCORES AVERAGED FOR EACH GROUP AT EACH STAGE OF THE EXPERIMENT

Perseverative Errors

| Group | Stage | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Transfer |
| *SSS* | — | 2.64 | 1.92 | 1.64 | 1.24 | 1.00 | — |
| *SUS* | — | 3.72 | 3.24 | 2.68 | 1.20 | 1.88 | — |
| *USS* | — | 3.80 | 2.16 | 3.68 | 1.84 | 1.76 | — |
| *UUS* | — | 2.56 | 1.92 | 2.64 | 1.56 | 2.00 | — |
| *SSU* | — | 3.88 | 1.20 | 3.36 | 2.24 | 1.16 | — |
| *SUU* | — | 5.96 | 6.12 | 3.80 | 3.20 | 2.36 | — |
| *USU* | — | 1.92 | 2.92 | 1.92 | 1.32 | 0.80 | — |
| *UUU* | — | 2.28 | 4.12 | 1.96 | 1.88 | 2.12 | — |

Non-Perseverative Errors

| Group | Stage | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Transfer |
| *SSS* | 3.40 | 1.76 | 1.12 | 0.84 | 1.36 | 0.64 | 0.04 |
| *SUS* | 3.88 | 2.20 | 1.20 | 1.16 | 1.28 | 1.08 | 0.76 |
| *USS* | 3.08 | 2.72 | 0.72 | 2.24 | 1.84 | 0.60 | 0.24 |
| *UUS* | 2.76 | 0.64 | 1.04 | 2.16 | 1.24 | 1.28 | 0.80 |
| *SSU* | 4.20 | 1.92 | 0.40 | 1.04 | 2.04 | 0.96 | 1.32 |
| *SUU* | 4.80 | 1.56 | 3.12 | 2.08 | 1.16 | 1.24 | 0.40 |
| *USU* | 3.76 | 0.56 | 1.32 | 1.36 | 0.80 | 0.84 | 1.76 |
| *UUU* | 3.32 | 1.20 | 1.92 | 1.00 | 2.56 | 1.40 | 0.40 |

Total Correct Responses

| Group | Stage | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Transfer |
| *SSS* | 0.16 | 2.44 | 2.16 | 2.20 | 3.84 | 0.76 | 0.00 |
| *SUS* | 1.48 | 5.00 | 2.36 | 2.24 | 2.32 | 1.56 | 0.52 |
| *USS* | 0.28 | 7.24 | 2.00 | 3.24 | 4.12 | 2.88 | 0.88 |
| *UUS* | 0.48 | 2.36 | 2.40 | 3.08 | 2.80 | 1.56 | 0.72 |
| *SSU* | 0.72 | 3.84 | 0.76 | 3.56 | 3.84 | 0.88 | 0.84 |
| *SUU* | 3.56 | 7.28 | 5.56 | 6.04 | 3.96 | 3.48 | 1.68 |
| *USU* | 0.52 | 2.52 | 1.48 | 3.92 | 3.32 | 1.36 | 1.88 |
| *UUU* | 0.72 | 2.52 | 2.52 | 1.44 | 3.80 | 2.24 | 0.76 |

[3] The initial perseverative error at each stage was eliminated in the scoring. Presumably this error serves to inform the *S* that a shift has taken place.

[4] Due to ambiguities inherent in the stimulus card–response card relationships correct responses frequently occur immediately after a shift before the first perseverative error is made. These were eliminated in scoring the number of correct responses at each stage.

data for the *U*-response cards. This has been done, and the results have been plotted in the three graphs of Fig. 2.

Examination of Fig. 2 reveals the general trend of the results. On stages 1, 2, 4, and 5, where the *color* and *form* categories are involved, there are relatively little differences introduced by the response cards. On the third, and to some extent on the sixth stage, the *number* stages, there are clear differences. The *S*-cards result in fewer perseverative and non-perseverative errors and require fewer correct responses before the criterion series is sorted. It would appear that constant configurational arrangement of the forms on the cards simplifies the number concept considerably.

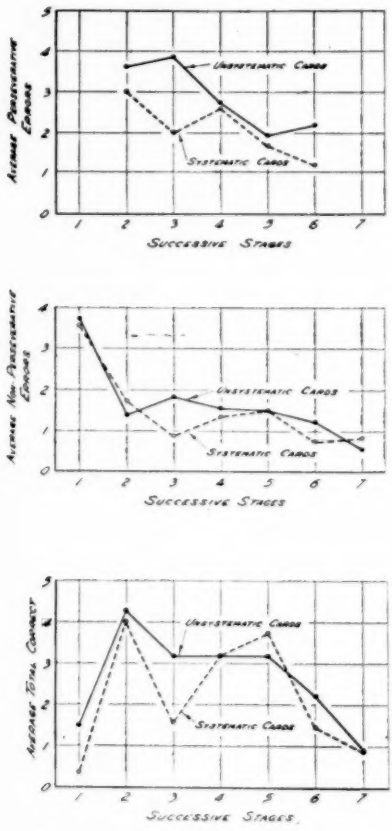


FIG. 2. Average number of perseverative errors, non-perseverative errors, and total correct responses graphed on successive stages for the systematic and unsystematic response cards

By regrouping the eight sets of data for each score in Table I in a standard manner (14, p. 42), the effects of the other variables and their interactions are obtained and may be graphed as in Fig. 2. Space does not permit this to be done here, and it will suffice to say that such neat, systematic results as those in Fig. 2 are not to be obtained for the other factors except on the transfer stage, which will be discussed later.

A complete statistical analysis of the data of Table I is summarized in Table II. An analysis of variance was performed for each score at each of the first six stages of the experiment. The *F*s from each analysis are given in a separate column of Table II, and at the foot of each column in parentheses is given the error mean-square, based upon replication, which was used to obtain the *F*s in the column above it.

In Table II most of the significant *F*s occur at the third stage, the initial number stage, of the experiment.[5] The effect of the difference in response cards is significant at the 1 per cent confidence level for all scores. Thus, the clear-cut results shown in Fig. 2 are statistically significant.

*Transfer effects.*—The data of Tables I and II and Fig. 1 show that a constant configuration simplifies the number response. If this is so it should also prove easier for Ss to

[5] Further research will be required to discover whether the other significant *F*s represent Type I errors or substantial characteristics of the Ss' behavior.

TABLE II

SUMMARIES OF ANALYSES OF VARIANCE OF DATA PRESENTED IN TABLE I ON STAGES 1 TO 6

*Fs* for Perseverative Errors

| Source of Variation | *df* | Stage | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Stimulus Cards | 1 | — | 5.68* | 1/2.97 | 1/3.47 | 1/2.15 | 1/48.12 |
| Response Cards | 1 | — | 1/1.08 | 9.45** | 1/24.71 | 1/2.45 | 3.44 |
| Transfer Cards | 1 | — | 1/3.21 | 4.78* | 1/35.58 | 2.22 | 1/92.54 |
| Stimulus × Response | 1 | — | 2.92 | 5.08* | 1.08 | 1/8.61 | 1/14.15 |
| Stimulus × Transfer | 1 | — | 5.68* | 1/8.58 | 4.90* | 2.60 | 1/1.76 |
| Response × Transfer | 1 | — | 1.21 | 4.63* | 1/24.71 | 1/1.04 | 1/1.96 |
| Stim × Resp × Trans | 1 | — | 1/15.93 | 1/1.18 | 6.12 | 1/137.75 | 1/6.76 |
| Error | 192 | — | *17.52* | *17.15* | *17.79* | *11.02* | *12.03* |

*Fs* for Non-Perseverative Errors

| Source of Variation | *df* | Stage | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Stimulus Cards | 1 | 1.17 | 2.25 | 1/2.60 | 1/1.09 | 1/3.70 | 1/25.77 |
| Response Cards | 1 | 1/93.97 | 1/1.29 | 7.53** | 1/3.45 | 1/32.15 | 3.58 |
| Transfer Cards | 1 | 1/1.10 | 1.81 | 3.91* | 1/3.45 | 1/1.89 | 1/1.52 |
| Stimulus × Response | 1 | 1/2.84 | 1/1.04 | 1.92 | 1.11 | 3.36 | 1/3.94 |
| Stimulus × Transfer | 1 | 1/41.76 | 1/1.91 | 1/22.96 | 3.41 | 1/16.72 | 1/67.00 |
| Response × Transfer | 1 | 0.00 | 1.54 | 4.64* | 1/14.98 | 1.82 | 1/13.40 |
| Stim × Resp × Trans | 1 | 1/167.00 | 5.18* | 3.03 | 1/2.95 | 7.46** | 1/30.45 |
| Error | 192 | *30.07* | *7.48* | *5.74* | *9.14* | *4.18* | *3.35* |

*Fs* for Total Correct

| Source of Variation | *df* | Stage | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Stimulus Cards | 1 | 4.17* | 1.37 | 1.03 | 1/1.51 | 1/1259. | 1/2.26 |
| Response Cards | 1 | 5.64* | 1/8.97 | 7.15** | 1/524.60 | 1/1.61 | 2.10 |
| Transfer Cards | 1 | 2.64 | 1/14.54 | 1/2.96 | 2.10 | 1/2.38 | 1/2.90 |
| Stimulus × Response | 1 | 3.84 | 10.52** | 2.19 | 3.17 | 1/25.69 | 3.53 |
| Stimulus × Transfer | 1 | 1.27 | 6.03* | 1/1.20 | 4.46* | 1/3.89 | 1.99 |
| Response × Transfer | 1 | 1/1.60 | 2.95 | 4.74* | 1/524.60 | 1.47 | 3.83 |
| Stim × Resp × Trans | 1 | 1/1.60 | 1.42 | 2.70 | 2.70 | 1/314.75 | 1/26.10 |
| Error | 192 | *11.52* | *35.18* | *18.12* | *26.33* | *25.18* | *13.05* |

* Significant at the 5 per cent confidence level.
** Significant at the 1 per cent confidence level.

transfer from *U*- to *S*-response cards than from *S*- to *U*-response cards, once they have learned to sort for *number*. The errors and correct responses on the seventh, or transfer, stage are totalled in the last column of Table I. Sub-grouping shows that transfer to *U*-cards is generally harder

TABLE III

SUMMARIES OF ANALYSES OF VARIANCE OF THE *Non-Perseverative Error Score* AND THE *Total Correct Score* AT THE *Transfer* STAGE OF THE EXPERIMENT

| Source of Variation | *df* | Non-Perseverative Errors | Total Correct |
|---|---|---|---|
| Response Cards | 1 | 1/1.05 | 1/288.50 |
| Stimulus Cards | 1 | 1/2.27 | 1/1.28 |
| *Transfer* | 1 | 3.95* | 5.01* |
| Stimulus-Response | 1 | 1/2.91 | 3.77 |
| Stimulus-Transfer | 1 | 1/25.31 | 1/2.00 |
| *Response-Transfer* | 1 | 12.04*** | 1/4.51 |
| Stimulus-Response-Transfer | 1 | 1/14.95 | 1/1.13 |
| Error Estimate | 192 | *3.29* | *5.77* |

* Significant at the 5 per cent confidence level.
*** Significant at the 0.1 per cent confidence level.

than transfer to *S*-cards and that actually shifting the deck of cards from *U*- to *S*-cards or from *S*- to *U*-cards results in greater difficulty than continued sorting with the same deck.

The analyses of variance of the transfer data are given in Table III, which was prepared in the same general manner as Table II. Table III shows that the nature of the response cards on the transfer stage introduces a significant difference in non-perseverative errors and in total correct responses. Actual shift of the deck of response cards produces a significant difference in non-perseverative errors, as is shown by the *F* for *Response-Transfer* interaction. Unfortunately, due to a defect in experimental procedure, by which decks were *actually shifted* only when the type of response card was to be shifted, this latter difference may be partly due to the slight confusion attendant upon handing the *S* a new deck of cards. The differences were, at any rate, in the predicted direction.

To find whether transferring from *U*- to *S*-cards was indeed easier than transferring from *S*- to *U*-cards, the non-perseverative error scores for the *SUS* and *UUS* groups were contrasted with the non-perseverative error scores for the *SSU* and *USU* groups. The average difference of 0.76 error favored the *SUS* and *UUS* groups, as was predicted above, and the *t* of 2.09 (192 *df*) was significant at the 5 per cent confidence level.

*Verbal reports.*—All quantitative data pointed to the greater difficulty of the *U*-cards, and the quantitative data were backed up by the verbal reports of the *S*s. At the end of each session, *S* was asked what he was doing and had been doing. Care was taken to avoid leading *S* in his replies. Thirty-eight of the 100 *S*s who sorted the *S*-response cards on stages 1 to 6 volunteered that they had sorted for color, form, and *configuration* rather than number. Some of these, but not all, recognized that there was a correspondence between configuration and number. One *S* of the 100 who had been using the *U*-cards claimed that he had been sorting according to configuration![6] The $\chi^2$ for the difference between these frequencies was 41.25, as corrected for continuity. For one *df* this $\chi^2$ is significant at the 0.1 per cent confidence level.

## SUMMARY AND DISCUSSION

The findings of the present study are clear. Sorting for number when the configuration of forms varied from card to card was more difficult than sorting for number when the configuration was constant. Apparently this difference in difficulty was most important when *S*s were first confronted

[6] This *S* may have been referring to the transfer stage of the experiment, where he sorted *S*-cards, or he may have been confused.

with number sorting, because the error and total correct scores did not differ significantly on the sixth (second *number*) stage of the test. Transfer to sorting *U*-cards from sorting *S*-cards was harder than the reverse. The verbal reports showed that over one-third of the *S*s made use of the constant configurations when that was possible. They apparently responded perceptually rather than analytically when they had the opportunity.

The first implication of these results is that a repetition of the previous experiment (**3**) on the relative difficulty of the number, form, and color concepts of the WCST using *U*-response cards might result in a changed order of difficulty. With the configurational aspects of the number concept removed, results more like those of Heidbreder should be obtained in spite of differences in procedure between the WCST and Heidbreder's "modified memory" and sorting experiments.

The broader implication of the results lies in their support of Heidbreder's hypothesis that *S*s respond perceptually more readily than analytically in abstraction situations. This hypothesis should be tested in a wide variety of situations, especially in the verbal realm, where it has already been shown that the ease of response to a concept is governed more by the clarity of linguistic expression than by the abstract complexity of the concept itself (**13**).

(Manuscript received
December 20, 1949)

## References

1. Berg, Esta A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15–22.
2. Grant, D. A., and Berg, Esta A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *J. exp. Psychol.*, 1948, **38**, 404–411.
3. Grant, D. A., Jones, O. E., and Tallantis, Billie. The relative difficulty of the number, form, and color concepts of a Weigl-type problem. *J. exp. Psychol.*, 1949, **39**, 552–557.
4. Heidbreder, Edna. Toward a dynamic psychology of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.
5. Heidbreder, Edna. The attainment of concepts: A psychological interpretation. *Trans. N. Y. Acad. Sci.*, 1945, **7**, 171–188.
6. Heidbreder, Edna. The attainment of concepts: I. Methodology and terminology. *J. gen. Psychol.*, 1946, **35**, 173–189.
7. Heidbreder, Edna. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35**, 191–223.
8. Heidbreder, Edna. The attainment of concepts: III. The process. *J. Psychol.*, 1947, **24**, 93–138.
9. Heidbreder, Edna, Bensley, Mary L., and Ivy, Margaret. The attainment of concepts: IV. Regularities and levels. *J. Psychol.*, 1948, **25**, 299–329.
10. Heidbreder, Edna, and Overstreet, Phoebe. The attainment of concepts: V. Critical features and contexts. *J. Psychol.*, 1948, **26**, 45–69.
11. Heidbreder, Edna. The attainment of concepts: VI. Exploratory experiments on conceptualization at perceptual levels. *J. Psychol.*, 1948, **26**, 193–216.
12. Heidbreder, Edna. The attainment of concepts: VII. Conceptual achievements during card-sorting. *J. Psychol.*, 1949, **27**, 3–39.
13. Heidbreder, Edna. The attainment of concepts: VIII. The conceptualization of verbally indicated instances. *J. Psychol.*, 1949, **27**, 263–309.
14. Snedecor, G. W. *Statistical methods*. Ames, Ia.: Collegiate Press, 1946.

# TRANSPOSITION IN CHILDREN AS A FUNCTION OF AGE [1]

BY ELIZABETH ALBERTS AND DAVID EHRENFREUND

*Psychological Laboratory, The State College of Washington*

In the typical transposition experiment *Ss* are first trained to respond differentially to two stimuli (e.g., dark gray vs. black), and are then tested by being presented with two other stimuli in the same continuum (e.g., medium gray and dark gray). If in the test situation the *Ss* respond in a manner consistent with the original training, in this case to the lighter stimulus, they are said to have transposed. One of the most well-established experimental findings in this type of experiment is the inverse relationship between the amount of transposition and the distance between the test and training stimuli. (1, 2, 3, 4, 6.)

According to Spence (5, 6) this phenomenon can be deduced from conditioning principles. In the learning of the original discrimination the successive reinforcements have cumulatively strengthened the excitatory tendency of the correct cue to the response of approaching it, whereas successive non-reinforcements have resulted in the cumulative development of an opposing inhibitory tendency to the response of approaching the incorrect stimulus cue. In order to predict which member of a new stimulus pair will be chosen in the test for transposition, Spence calculates the relative strengths of their respective excitatory tendencies from the following learning principles: (1) generalization of the above-mentioned excitatory tendency to other stimuli along some continuum of physical properties; (2) generalization of inhibitory tendencies to other stimuli along some such continuum. As a result of these two opposing effects, any stimulus in the continuum will have a certain amount of effective excitation depending on the algebraic summation of inhibition and excitation.

A clear statement of Spence's hypothesis can best be presented by the reproduction of a diagram from one of his articles (7), in which he has illustrated the results of the interaction of generalization of excitation and inhibition in a hypothetical situation involving size.

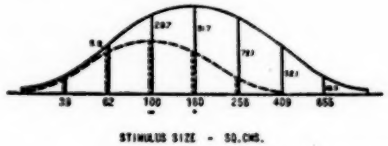


FIG. 1. Hypothetical generalization curves of excitation and inhibition as used by Spence in his theory of transposition

Fig. 1 represents a size discrimination problem in which the *S* has already learned to respond to 160 vs. 100. The solid line represents the curve of generalization of excitation built up to 160, while the broken line represents the curve of generalization of inhibition to stimulus 100. These two summate algebraically so that the effective excitatory strength of any stimulus in the size continuum is equal to the difference between the two and is represented by the upper set of numbers in the diagram. Thus, it is evident that *Ss* trained to choose the larger of the stimulus pair, 100 vs. 160,

[1] E. Alberts is now at the University of Wisconsin. This study constitutes the major portion of a thesis submitted by E. Alberts to the State College of Washington for the degree of Master of Science, 1948. D. Ehrenfreund served as thesis advisor.

will respond to the larger of a new pair, 160 vs. 256, but will fail to transpose the original learning as the difference between the training pair and the test pair increases.

Furthermore, it is evident that on transposition tests one step away from the training pair, the stimulus which originally was positive will be reacted to as if it were negative. However, all theories of transposition predict this result under certain circumstances. Studies demonstrating this fact, therefore, do not necessarily constitute a test between theories based on Gestalt principles and those based on conditioning principles.

The first comprehensive experiment attempting to test Spence's theory using child *Ss* was recently reported by Kuenne (4). In this study 44 children, ranging in mental age from three to six, were divided into two matched groups. She trained both groups to choose the smaller stimulus of the largest of six pairs of stimuli. One group was then tested one step down, the other group, five steps down. Subsequently, both groups were again brought to criterion and each tested on the pair opposite that of their previous test. It was found that MA is not correlated with transposition tests *one* step removed from the training pair, the median transposition scores of all four MA groups being very high. However, a relationship between MA and transposition tests *five* steps removed was clearly demonstrated. The product-moment correlation of .66 was significant well beyond the .1 per cent level. From these results and from data on verbalization, Kuenne concluded that young children pass by stages from control of behavior on a non-verbal level to verbal learning at a later age, and that, in general, Spence's theory had been confirmed.

## Statement of the Problem

The present investigation takes its point of departure from the study by Kuenne cited above. In that experiment the application of conditioning theory to transposition behavior of children was based on the hypothesis that

> the mechanisms assumed by Spence to underlie this behavior in animals are also operative in the young child, and that with the development in older children of the capacity to employ verbal responses in such behavior situations, a shift occurs to the verbal type of control (4, p. 474).

As noted above, one of the implications of this hypothesis which Kuenne confirmed is that there is a direct functional relationship between mental age and the amount of transposition on stimuli five steps removed from the training pair.

Another experimentally testable implication of the same hypothesis, which this study attempts to confirm, is that children tested at increasingly distant steps from the training pair, i.e., one, two, three, four, and five steps away, should produce different gradients according to their mental ages. Thus it is expected that under these conditions the data for younger children will form a rather steep gradient whereas data for older children will form a horizontal gradient reflecting high transposition scores on all tests.

The design of the experiment followed, in general, that of Kendler (2) for transposition downward. The procedure followed, generally, that of Kuenne (4) with several exceptions. Two age groups were employed instead of four. Each *S* was trained and tested only once, rather than twice, and transposition was tested at one, two, four, and five steps, instead of at near and far points only.

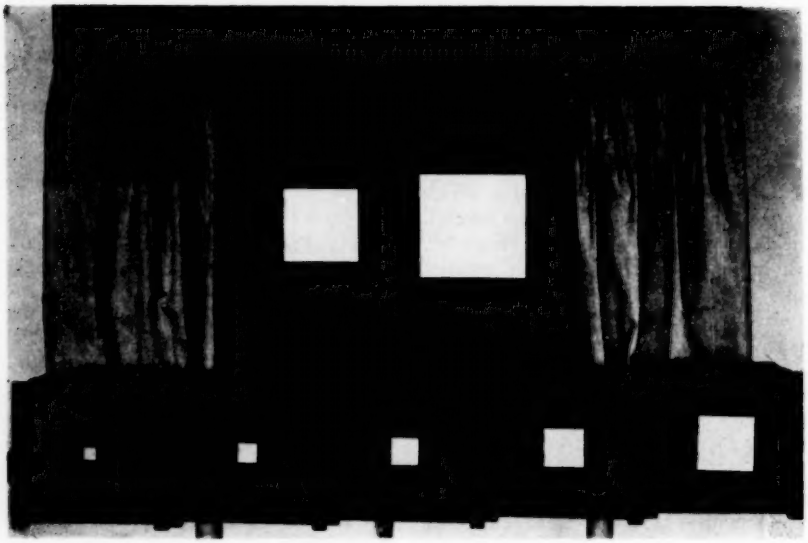

FIG. 2. Photograph of apparatus

## EXPERIMENTAL PROCEDURE

*Subjects.*—Forty-eight children from the nursery school of the Home Economics Department of the State College of Washington served as *Ss*. Eight children were dropped, leaving 40 for whom data are reported.[2]

Since mental age scores were unavailable, *Ss* were selected on the basis of chronological age. The younger group contained 22 *Ss*, ranging in age from three years to three years, eleven months. The older group contained 18 *Ss* ranging from four years, seven months, to five years, five months. Those children aged four to four and one-half were excluded in order to prevent excessive overlap in development between the two groups.

*Apparatus.*—The apparatus, shown in Fig. 2, consisted of a panel and seven boxes. The panel is shown on a table for purposes of photography, but when in use it stood on the floor. Before the panel are shown the various boxes used as test stimuli. The panel, 4 ft. by 6 ft., was cut from fiber-board. Two square openings, each 14 in. on a side, were cut near the center of the panel 2 in. apart. Reinforcing boards were added on the front and back so that the whole apparatus stood firmly by itself. The entire front was painted flat black. A string stretching from side to side near the top supported two opaque curtains which could be drawn closed to conceal the openings from *S*.

Each of the seven boxes measured 14 in. on a side, and could be inserted into either of the apertures in the panel and held there by a cupboard latch on the back of the panel. There were no tops on the boxes, permitting *E* to place the reward therein, silently and unseen. The front of each box was painted black and fitted with a white door which served as the stimulus. The doors varied in size from 1.4 in. on a side to 11.3 in. on a side; the ratio of the area of one door to the next was 2:1. The stimuli were paired to form conditions according to the following diagram.

| A | B | C | D | E | F | |
|---|---|---|---|---|---|---|
| 128 | 64 | 32 | 16 | 8 | 4 | 2 |

The numbers refer to area in square in. Condition A was the training pair (64 positive vs. 128 negative). Conditions B through F, with the exception of D, were the pairs used in transposition tests.

When used as a positive stimulus in the training series, or as a stimulus in the test series, the

[2] Seven children were dropped from the experiment for such reasons as illness, withdrawal from school, or refusal to continue. Data for one subject were excluded because he had been responding to extraneous cues.

floor of a box was lined with white paper. When used as the negative stimulus of the training pair it was lined with black paper. It was thought that this device, and the fact that *S*s operated the actual stimuli, would facilitate learning.

*Training series.*—The procedure during the training series was the same for both groups. Each *S* was asked to participate in a game involving the finding and collecting of gumdrops. If cooperative, he was given a toy dog to carry to the laboratory and to play with between trials. If he was unwilling to go, another child was asked.

Upon entering the experimental room, *S* was seated in a small chair about 3 ft. away from the center of the panel. *E* then drew the curtains apart, pointed out the doors forming condition A, and demonstrated how they could be opened by pulling each about ½ in. ajar and closing them immediately. It was explained that *E* would hide a gumdrop behind one of the doors, and that *S* should open one of them to find it.

When a trial was successful (i.e., *S* opened the smaller door), *S* took out the gumdrop and placed it upon a napkin which had been spread on another small chair near by, while *E* closed the door and the curtains.

For an unsuccessful trial the procedure was as follows: Immediately after *S* opened the wrong door, *E* pulled the curtain over the cor-

TABLE I

LEARNING AND TEST DATA FOR THE YOUNGER GROUP

| Subjects | CA Mos. | No. of Training Trials | Response to 1st Test Trial | % "Smaller" Responses to Test Trials 1–5 | % "Smaller" Responses to Test Trials 1–10 |
|---|---|---|---|---|---|
| Subgroup I-B | | | | | |
| 1 | 40 | 10 | + | 100 | 100 |
| 2 | 47 | 39 | + | 100 | 100 |
| 3 | 47 | 27 | + | 100 | 100 |
| 4 | 38 | 27 | + | 100 | 100 |
| Mean | 43.0 | 25.8 | | 100.0 | 100.0 |
| Median | 47.0 | 27.0 | | 100.0 | 100.0 |
| Subgroup I-C | | | | | |
| 5 | 40 | 11 | − | 80 | 90 |
| 6 | 44 | 25 | + | 100 | 100 |
| 7 | 45 | 12 | + | 60 | 50 |
| 8 | 45 | 80 | − | 80 | 90 |
| 9 | 39 | 40 | − | 40 | 50 |
| 10 | 37 | 55 | − | 80 | 90 |
| Mean | 41.7 | 37.2 | | 73.3 | 78.3 |
| Median | 42.0 | 32.5 | | 80.0 | 90.0 |
| Subgroup I-E | | | | | |
| 11 | 44 | 70 | + | 40 | 60 |
| 12 | 37 | 25 | − | 80 | 80 |
| 13 | 38 | 14 | − | 20 | 40 |
| 14 | 38 | 55 | + | 60 | 60 |
| 15 | 44 | 125 | + | 60 | 70 |
| 16 | 38 | 210 | − | 60 | 50 |
| Mean | 39.8 | 83.2 | | 53.3 | 60.0 |
| Median | 38.0 | 62.5 | | 60.0 | 60.0 |
| Subgroup I-F | | | | | |
| 17 | 40 | 12 | − | 20 | 30 |
| 18 | 42 | 12 | − | 40 | 50 |
| 19 | 44 | 10 | − | 80 | 90 |
| 20 | 47 | 11 | + | 40 | 50 |
| 21 | 39 | 25 | + | 60 | 60 |
| 22 | 41 | 194 | + | 60 | 60 |
| Mean | 42.2 | 44.0 | | 50.0 | 56.7 |
| Median | 41.5 | 12.0 | | 50.0 | 55.0 |

TABLE II

LEARNING AND TEST DATA FOR THE OLDER GROUP

| Subjects | CA Mos. | No. of Training Trials | Response to 1st Test Trial | % "Smaller" Responses to Test Trials 1–5 | % "Smaller" Responses to Test Trials 1–10 |
|---|---|---|---|---|---|
| Subgroup II-C | | | | | |
| 23 | 54 | 10 | + | 80 | 90 |
| 24 | 62 | 11 | + | 100 | 100 |
| 25 | 58 | 25 | + | 100 | 100 |
| 26 | 54 | 15 | + | 100 | 100 |
| 27 | 54 | 164 | + | 80 | 70 |
| 28 | 60 | 10 | + | 100 | 100 |
| Mean | 57.0 | 39.2 | | 93.3 | 93.3 |
| Median | 56.0 | 13.0 | | 100.0 | 100.0 |
| Subgroup II-E | | | | | |
| 29 | 56 | 35 | − | 80 | 90 |
| 30 | 65 | 27 | + | 100 | 100 |
| 31 | 57 | 25 | − | 80 | 80 |
| 32 | 57 | 22 | − | 80 | 90 |
| 33 | 61 | 30 | + | 60 | 60 |
| 34 | 58 | 59 | − | 80 | 90 |
| Mean | 59.0 | 33.0 | | 80.0 | 85.0 |
| Median | 57.5 | 28.5 | | 80.0 | 90.0 |
| Subgroup II-F | | | | | |
| 35 | 55 | 13 | + | 100 | 100 |
| 36 | 59 | 47 | + | 100 | 100 |
| 37 | 58 | 12 | + | 100 | 100 |
| 38 | 62 | 15 | + | 100 | 100 |
| 39 | 63 | 10 | − | 60 | 70 |
| 40 | 61 | 13 | + | 100 | 100 |
| Mean | 59.7 | 18.3 | | 93.3 | 95.0 |
| Median | 60.0 | 13.0 | | 100.0 | 100.0 |

rect door to prevent *S* from opening it in turn. *S* was then assured of another opportunity to find the candy, and was reseated. The first door was closed, and the curtain was pulled across it.

Between trials, after both curtains had been drawn closed, *E*, who remained behind the panel for the duration of the session for each child, recorded the data for the previous trial. The response, the time (measured from the opening of the curtains by *E* to the opening of the door by *S*) and any verbalization in which *S* mentioned the stimuli were entered on a mimeographed data sheet. Also, the boxes were taken out and replaced after each trial whether or not they were interchanged. The gumdrop was then put in proper position for the next trial.

Fifteen trials were presented on each school day. Training on condition A continued until *S* reached the learning criterion of any nine correct choices in ten consecutive trials.

*Test series.*—On the day following the criterion, each *S* was given a transposition test of ten trials, during which all choices were rewarded. Transposition was tested at one step, condition B, with four younger children. These *S*s will be referred to as subgroup I-B. Conditions C, E, and F each were tested with six children from each age group. These subgroups of younger children will be referred to as I-C, I-E, and I-F, and the subgroups of older children, as II-C, II-E, and II-F. The *S*s were assigned to the several subgroups at random prior to their introduction to the experimental situation.

*Verbalization.*—After completing the test series, if *S* had not verbalized size as the cue, he was again shown the training pair and asked how he made a correct choice. The first *S*s trained (nine younger, one older) were not questioned after the last trial since it was originally planned to retrain and test them at other points. Due to lack of time this plan was abandoned and thereafter all *S*s were questioned after the last test trial. All replies were recorded.

## Results and Discussion

*Training.*—The data collected for the 40 *Ss* are shown in Tables I and II. The third column lists the number of trials required by each *S* to reach the criterion of learning. Although there are large individual differences within both age groups, the mean number of trials required by older children, 30.2, was considerably lower than the mean number required by younger children, 54.8. These data support Kuenne's finding that older children learn a discrimination problem faster than do younger children.

A comparison of the mean and median numbers of training trials received by the subgroups within each age group reveals rather large differences. Thus, the practice of randomly assigning *Ss* to the various subgroups has resulted in considerable inequality between subgroups in speed of learning. However, rank order correlations between speed of learning and transposition scores for the several subgroups were distributed unsystematically between −.51 and +.40. This was also to be expected on the basis of Kuenne's data where it was found that little or no relationship obtained between number of trials required for learning and transposition score.

*Transposition.*—In the fourth column of the tables, a plus sign indicates that on the first test trial, response was to the "correct," or smaller stimulus. A minus sign indicates the reverse. It is generally agreed that response to the first test trial is one of the most valid measures of transposition. However, due to the small number of *Ss*, the response to the first test trial in this study is probably unstable.

Mean scores on the first five test trials, presented in column 5, are more

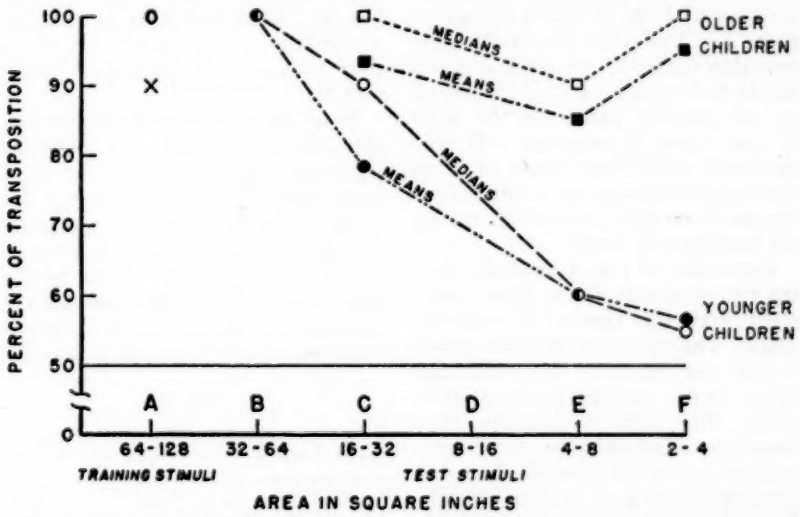


Fig. 3. Percentage of transposition obtained by each of the subgroups at the various steps tested. The point at X is the criterion score reached by all *Ss* on the training stimuli. The score at O is the assumed transposition score of a group had it been tested on the training stimuli.

consistent, but show a great deal of individual variation within subgroups. For the younger children, scores on far tests range down to 20 per cent, suggesting the reversals of response which are implied by the Spence hypothesis. The mean scores of the ten test trials, in column 6, agree with those for the first five test trials and show the same trends.

The trends for the ten test trials are summarized graphically in Fig. 3, in which transposition scores are plotted as a function of the distance of the test pair from the training pair.

No children from the older group were tested on Condition B, for it was assumed that if older children made high transposition scores at points more than one step away, they would, if tested at one step, have made scores at least as high. For this reason, the gradients for younger children begin farther to the left than do those of the older children.

The criterion score, reached on the training trials by all *S*s, is plotted at point X. The point at 0 is the most probable transposition score a group would have made had it been tested on the training pair. On the basis of the scores of subgroup I-B (100 per cent), which was tested one step from the training pair, it seemed unnecessary actually to test a group on the training pair itself.

Inspection of Fig. 3 disclosed that the two gradients are in close agreement with the theoretical expectations. The curve for the older group is high and relatively flat, while the curve for the younger group is fairly steep. The former indicates that there is no relationship between transposition scores for older children and distance of test from training stimuli. The high, fairly horizontal curve is that which is expected when responses on transposition tests are verbally controlled. The sloping gradient for the younger group, however, suggests an inverse relationship between the variables plotted. Furthermore, it is in general the form of gradient expected when responses to several distant test stimuli are based on those mechanisms assumed by Spence to be operative in pre-verbal subjects.

Because of their form, the two curves draw farther apart at D, E, and F, as transposition was tested at more distant points. This disparity between gradients further supports the belief in the difference in mode of response between the two groups. These findings are in close accord with Kuenne's results (4). As in her study, both groups had high scores on the near test but differed greatly on the far test.

Finally, some attention should be given to the low point in the gradient of the older group, which was contrary to expectations. Data for subgroup II-E show that the drop in the mean score was due mainly to one *S*, No. 33. The use of more *S*s probably would have counteracted this effect, thus producing a horizontal curve with less variability.

In order to demonstrate the significance of the data, the following statistical analyses were made: (1) a *t*-test between the mean scores for trials on the near and far transposition tests of the younger children; (2) a *t*-test between the far scores for the two groups. Table III presents the statistical evidence for the difference between scores at conditions B and C

TABLE III

DIFFERENCE BETWEEN CONDITIONS B AND C, AND E AND F FOR THE YOUNGER GROUP

| Condition | Mean | *S.D.* | Diff. | *t* | *P* |
|---|---|---|---|---|---|
| B and C | 8.70 | 1.90 | 2.87 | 3.69 | .01 |
| E and F | 5.83 | 1.58 | | | |

TABLE IV

DIFFERENCE BETWEEN OLDER AND YOUNGER GROUPS ON CONDITIONS E AND F

| Group | Mean | S.D. | Diff. | t | P |
|---|---|---|---|---|---|
| Younger | 5.83 | 1.58 | 3.17 | 5.14 | .001 |
| Older | 9.00 | 1.29 | | | |

and those for conditions E and F in the younger groups. As indicated in Table III, the mean transposition score for tests (B and C) near the training pair is significantly different from the mean transposition score for tsets (E and F) farther removed from the training pair. The obtained *t* value of 3.69 far exceeds the 2.84 required for the 1 per cent level of confidence. We can therefore make the statement with a high degree of probability that the gradient for the younger *Ss* is not horizontal.

Similar data are presented in Table IV as statistical evidence for the difference between the scores of the older and younger groups at conditions E and F. Table IV indicates that the null hypothesis can be rejected at the .1 per cent level of confidence.

*Latency.*—Mean latency scores were computed for each subgroup on both training trials and test trials. There was very little variation between age groups or between training and test trials, since the extremes varied by only 1.53 sec. There was a tendency, however, for the latency of the first test trial to exceed that of the other combinations.

*Verbalization.*—An analysis of the verbal responses revealed that they could be differentiated according to the following criteria: (1) spontaneous or elicited, and (2) reference to size without its significance to the solution or verbalizing the principle of solution. Therefore, *Ss* were classified in one of the following four categories: (A) elicited size; (B) spontaneous size; (C) elicited principle; (D) spontaneous principle. These categories are mutually exclusive in that no *Ss* were checked in more than one category. For example, if an *S* spontaneously mentioned size but not its significance to the solution (category B), and then in response to questioning following the test trials, verbalized the principle (category C), he was placed in the latter category and not in the former. The similarity between our categories and those of Kuenne are at once apparent. In both instances, categories A and B are non-generalizing, while C and D contain *Ss* that made verbal generalizations.

Table V represents the category in the first column, the number and per cent of *Ss* in each category for the older and younger groups respectively in the next four columns. Categories A and B have been placed together, since they contain all *Ss* who failed to verbalize the principle involved in learning the discrimination; C and D present data for those who, spontaneously or in response to questioning, verbalized the principle.

TABLE V

SUBJECTS CLASSIFIED ACCORDING TO VERBALIZATION CATEGORIES *

| Category | Older | | Younger | |
|---|---|---|---|---|
| | *N* | % | *N* | % |
| A | 0 | 0 | 5 | 31.25 |
| B | 2 | 11.1 | 5 | 31.25 |
| C | 9 | 50.0 | 1 | 6.25 |
| D | 7 | 38.9 | 5 | 31.25 |

* Of the *Ss* (nine younger, one older) previously mentioned as not being questioned after the last test trial, the table includes the one older and three of the younger since they spontaneously verbalized the principle. Of the six younger *Ss* not included, five spontaneously mentioned size.

Due to the incompleteness of the data, any conclusions are necessarily only tentative. With these reservations, we note that older *Ss* tend to be in the generalizing categories, (88.9 per cent) while a majority of younger *Ss* (62.5 per cent) are in the non-generalizing groups. That a large minority of younger *Ss*, (37.5 per cent) are classified as having verbalized the principle, may be due in part to the use of chronological age instead of mental age and in part to the previously mentioned incompleteness of the data.

## Summary and Conclusions

The present experiment was based on the hypothesis that those mechanisms indicated by the results of conditioning experiments to be operative in the transposition behavior of infrahuman *Ss*, are also operative in the transposition behavior of young children. Contrariwise, it is postulated that the transposition behavior of older children is verbally controlled. This experiment attempted to test empirically the implications of the hypothesis: namely, that young children tested at several steps from the training stimuli would produce a steep gradient, while older children similarly tested would produce a horizontal gradient.

Forty children were divided into two groups, the first ranging in age from 3–0 to 3–11, the second from 4–5 to 5–6. All were trained to choose the smaller of a pair of squares, 64 vs. 128. Then *Ss* from each group were tested at points one, two, four, and five steps from the original stimuli. Spontaneous verbalizations were recorded, and a statement of the solution was solicited from most *Ss* at the close of experimentation.

From an analysis of the results the followni conclusions were drawn:

1. The gradient for the older group is horizontal at a high level of transposition score.

2. The gradient for the younger group drops from a high to a low transposition score as test stimuli are removed from training stimuli.

3. An inspection of the curves indicates that the departure of the curve for the younger group from that of a horizontal one is of the general form and direction which is predicted in our original hypothesis.

4. Tentatively confirming a previous investigation by Kuenne, the data on verbalization indicated that older children tend to verbalize their responses to size as the cue aspect, whereas younger children do not.

5. The various data (especially 3 above) lend striking confirmation to the theory of transposition advanced by Spence.

(Manuscript received
January 17, 1950)

## References

1. Gulliksen, H. Studies of transfer of response: I. Relative versus absolute factors in the discrimination of size by the white rat. *J. Psychol.*, 1932, 40, 37–51.
2. Kendler, Sylvia T. S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, 40, 552–562.
3. Kluver, H. *Behavior mechanisms in monkeys.* Chicago: Univ. of Chicago Press, 1933.
4. Kuenne, Margaret R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, 36, 471–490.
5. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, 43, 427–449.
6. Spence, K. W. Analysis of formation of visual discrimination habits in chimpanzee. *J. comp. Psychol.*, 1937, 23, 77–100.
7. Spence, K. W. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, 31, 257–271.

# AN EXPERIMENTAL INVESTIGATION OF REACTIVE INHIBITION AND CONDITIONED INHIBITION

BY KAY C. MONTGOMERY[1]

*Harvard University*[2]

The relation between the amount of effort involved in a given response and the processes of acquisition and extinction of that response represents a problem to which an increasing amount of research has been devoted in recent years. In general it appears that greater amounts of work result in slower learning and in more rapid extinction (4, 10, 11, 13). When the relative spacing of responses is viewed as a temporal distribution of work, the problem of massed versus spaced learning becomes another aspect of the role of effort (16). Under a wide variety of conditions, closely spaced trials retard learning and facilitate extinction (8, 12). Thus, although the evidence is not in complete agreement, it seems that the effect of increasing work, either by requiring a more effortful response or by requiring more responses per unit time, is a reduction in learning rate and an increase in extinction rate.

A number of writers have concerned themselves with the explanation of the role of effort in learning (16). The most comprehensive and systematic of these explanations is that proposed by Hull in his *Principles of Behavior* (9). In Hull's system the two postulates most pertinent to the present problem are VIII and IX. Postulate VIII states in effect that whenever a reaction is evoked in an organism there is created a negative drive state, reactive inhibition ($I_r$), that possesses the capacity to inhibit the repetition of that response (9, p. 300). $I_r$ is said to be a positively accelerated increasing function of the amount of work involved in the response and to dissipate spontaneously as a simple decay function of time. Postulate IX asserts that stimuli closely associated in time with the cessation of a response become conditioned to the reactive inhibition generated by the response, thereby setting up conditioned inhibition ($_sI_r$) (9, p. 300). Essentially what the $_sI_r$ construct denotes is a negative habit of *not-responding* that acts opposite in direction to the positive habit of responding and that possesses all of the characteristics exhibited by positive habits, e.g., stimulus generalization. Postulate IX further asserts that $_sI_r$ summates with $I_r$ to produce a net inhibitory potential ($I_r$) and that response strength is a function of the total reaction potential ($_sE_r$) minus $I_r$, i.e., a function of "effective reaction potential" ($_s\bar{E}_r$). These two postulates, when taken together with the others in Hull's system, make possible the prediction of not only the differential effects of work on learning, but also a wide range of other behavioral phenomena such as reminiscence, spontaneous recovery, and extinction.

In view of their explanatory power, it seems desirable to subject these two postulates to experimental confirmation. However, because both involve hypothetical constructs, they cannot be tested directly. It is necessary to deduce from them, in conjunction with other postulates in the system, theorems which then may be experimentally tested. Several such studies have been carried out in the past few years (e.g., 2, 15, 17). Although the results have in general been favorable, it should be pointed out that in

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. at the University of Chicago. The author wishes to express his gratitude to Dr. W. D. Neff.

[2] Now at Cornell University.

only one of these investigations was an attempt made to vary simultaneously in the same experimental situation the two important independent variables (amount of work and inter-trial interval) specified in the postulates. Kimble and Bilodeau (11), studying motor learning in human subjects, investigated performance as a function of *length of work period* and length of rest period. Thus, amount of work was varied by employing different lengths of *continuous* work. Another approach to the problem is to study the effects on performance of these two variables in a situation where amount of work does *not* depend upon length of work period. This involves the selection of a relatively specific response or response sequence that can be evoked at any desired time and in which the amount of work required can be varied independently of time. The present experiment represents such an attempt.

The purpose of the present experiment, then, is to test the validity of certain theorems deduced from Hull's Postulates VIII and IX in a situation where the two independent variables specified in the postulates are varied simultaneously. The experimental situation chosen was based on the following considerations: (1) the task had to be one involving a relatively specific response sequence that could be evoked at any desired time; (2) the task had to be one in which *amount of work per response* could be conveniently varied without introducing such undesired variables as length of work period or delay in reinforcement; (3) the task had to be simple enough to reveal the unclouded operation of the independent variables.[3] A combination maze-running and lever-pressing situation, much the same as that employed by Thompson (17), was selected. The response sequence was identifiable—running to either arm of a single-unit Y-maze and pressing a lever—and could be evoked when desired; work was easily varied by counter-balancing the levers with any desired amount of weight; alternation was possible; and the situation was of such a nature that any effects produced by the two independent variables would clearly manifest themselves.

[3] A second purpose of the experiment was to study the phenomenon of spontaneous alternation as a function of work and inter-trial interval. The results of this part of the investigation will be reported in a later paper.

## Experimental Design

A simple factorial design was chosen in order to combine three values of work—15, 40, and 90 gm. of bar-weight—with three of inter-trial interval—20, 45, and 90 sec. Five albino rats were assigned at random to each of the nine experimental groups. The measure of performance during acquisition and extinction was response time.

## Procedure

*Subjects.*—Forty-five male albino rats of the Fairfield strain were used. The animals were 60–70 days old at the beginning of the experiment; their initial weights ranged from 165–185 gm.; and all were experimentally naive.

*Apparatus.*—Figure 1 is a diagram of the apparatus and of the experimental arrangement. The maze was constructed from ¾-in. white pine and was painted flat black except for the ends of the two goal boxes, which were flat white. All alleys were 4 in. high; each had a hinged wire mesh top. The one-way doors were transparent and were counter-balanced for easy movement. The levers, constructed from brass rods ¼ in. in diameter, were 3 in. long and 3 in. from the floor. When *S* left the starting stem and went through either door, an electric chronoscope was started; when *S* moved either lever a certain distance downward (the same for both levers and for all *S*s), the chronoscope stopped, a food pellet was delivered, and the apparatus was automatically set for the next trial. Thus the time taken by *S* to press the lever after entering a goal arm could be measured precisely. Care was taken to insure that the door switches were adjusted in such a manner as to provide a relatively standard measure of response time, i.e., that on the average the clock would start when *S* was about halfway through the door.

*Pre-training.*—Pre-training lasted two days. On the first day each *S* was placed in a small training box containing a lever (immovable) and a food tray. After a 2-min. exploratory period

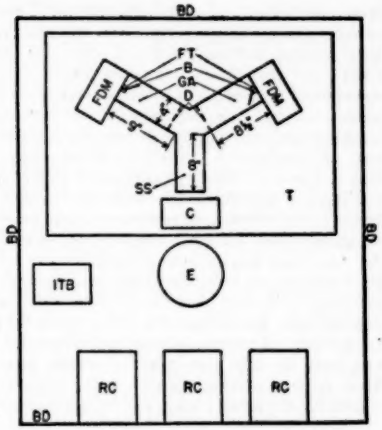


FIG. 1. Diagram of the experimental arrangement and apparatus. SS is the starting stem of the maze, D a one-way door, GA a goal arm, B a lever, FT a food tray, FDM a food-delivery mechanism, and C the chronoscope. The apparatus was placed on a table T in front of which sat the experimenter *E*. The subjects were placed in box ITB between trials and in cages RC between blocks of trials. Burlap drapes BD enclosed the apparatus on all sides. A single 20-watt bulb (not shown in diagram) was suspended directly over the choice-point of the maze.

each animal received ten pellets; these were delivered silently, one at a time. The *S* was then removed and after a period of at least 10 min. was returned to the box. *E* then delivered 15 more pellets, this time by operating the food device. On the second day each *S* received 35 reinforced bar-pressing trials. The rats were given ten trials during which the bar was counter-weighted with 5 gm., followed by five trials with each of the following weights: 15, 30, 50, 75, and 95 gm. Reinforcement was contingent upon the *S*'s depressing the lever a certain distance downward. At least 5 min. intervened between the trial-blocks for different weights. Response latencies were recorded with the aid of a stop watch. Thus at the completion of the pre-training period each animal had received 35 reinforced trials and had learned to press the bar when it was counter-weighted with 95 gm.

*Acquisition.*—The acquisition series consisted of 100 trials in the Y-maze, 20 trials on each of five successive days. On day 1 the animals received their first experience in the maze. Each *S* was first fed three pellets in each goal box. Immediately thereafter the first trial was given: the rat was placed by hand in the stem of the Y and allowed to run to either arm, press the lever, and eat for 5 sec. The animal was then removed from the maze and placed in a small box at one side of *E*. After the appropriate inter-trial interval this procedure was repeated. Ten such trials were given in succession, and then *S* was placed in a cage at the other end of the room. After a rest period of at least 10 min., usually 15, another block of ten trials was given.

The side to which *S* went (R or L) and the response time (time between entering a goal box and depressing the lever the required distance) were recorded for each run. The animals were given their daily ration immediately after the experimental session. The procedure for the succeeding four days was the same as on the first day with one exception: the rats were not fed three pellets in each goal arm before their initial trial of the day. Thus at the completion of the acquisition series each animal had received 100 reinforced trials in the maze, ten blocks of ten trials each.

*Extinction.*—Two days of extinction trials were given. The procedure was the same as for the acquisition series with the following exceptions: (1) the 20 trials per day were given in succession, not in two blocks of ten each; (2) the food reward was omitted; and (3) the food devices were disconnected so that the secondary reinforcement provided by their characteristic noise was eliminated. The same measures were recorded. Originally three or four days of extinction trials were planned; however, not only were response times excessively variable as early as the second day, but also after about the thirtieth to fortieth trial it proved extremely difficult if not impossible to control the inter-trial interval. Some *S*s would remain in the starting stem for periods exceeding their respective inter-trial intervals; others did not hesitate to enter a goal box. It was decided, therefore, to continue extinction for only two days.

*Control of variables pertaining to the subjects.*—The factors to be considered here are individual differences in motivation, body weight, activity level, and emotion. A feeding rhythm was established by keeping each *S* on a maintenance diet of 10 gm. of laboratory chow per day for at least five successive days before pre-training was begun. Except for certain minor adjustments to be noted later, the total amount of food per day throughout the experiment was held constant at 10 gm. At the beginning of each experimental session all animals had been deprived of food for about 21 to 22 hr.

Body weight was controlled in the following manner: From the available rats a relatively homogeneous weight group was chosen. Each

of these *S*s was then assigned at random to one of the nine experimental groups. The animals were weighed daily immediately before the experimental session and their daily rations adjusted slightly from time to time in such a manner as to maintain a constant body weight for each rat. Analysis of the data concerned with body weights indicated that the various groups were homogeneous. Because all *S*s were of the same age, sex, and strain, it is assumed that the procedure of random assignment equalized the factor of activity level. The fact that motivation was controlled makes this presumption all the more likely.

Random assignment insured at least partial control of emotionality. In addition, data on emotionality were collected by recording measures of urination and defecation for each *S* during the pre-training period.[4] *E* judged the amount of urination to be "none," "slight," "medium," or "great." These judgments were converted to a scale ranging from 0 to 3. Defecation was measured by counting the number of feces. Analysis of these data yielded no evidence of differences among groups. During the course of the experiment an effort was made to equalize handling and eating times for all animals.

*Control of apparatus variables.*—Intra-maze visual stimuli were controlled by making the two goal boxes as nearly identical as possible. Extra-maze visual cues were made symmetrical in two ways: (1) a homogeneous extra-maze visual background was provided by draping burlap on all sides of the apparatus; and (2) the experiment was conducted in a semi-darkened room, the only light being a 20-watt bulb hung directly over the choice-point. Auditory cues were of three kinds: the sound of the chronoscope, the characteristic noise of the food devices, and extraneous stimuli. The clock was placed directly in front of the stem of the Y so that its sound was symmetrical to the goal boxes. The sounds of the food mechanisms were made equivalent as far as possible. Because the experiment was conducted in a sound-proof room extraneous auditory stimuli were at a minimum; moreover, *E* sat directly in front of the Y-stem and placed the rats not being run, behind him. Olfactory cues were equalized by placing food in both food devices on every acquisition trial, and by completely eliminating food during extinction.

*Control of task variables.*—The task consisted of entering either arm of the maze, pressing a lever, and receiving and eating a small pellet of food. To insure that the responses were roughly equivalent on both sides of the maze, the levers were oriented in such a fashion that the *S* was forced to turn right in the goal box to press the lever whether it chose the R or L arm. On the basis of preliminary experimentation the amounts of force required for food delivery were chosen so that the responses for the various work groups were qualitatively not greatly different. *E* attempted to control the initial orientation of *S* in the Y-stem by placing each animal in the starting box alternately with the R and L hands. The amount of reward, a single pellet weighing about 0.1 gm., was constant for all *S*s. An effort was made to prevent the animals from pressing the lever more than once per trial, provided, of course, that the amplitude of the first response was sufficient to activate the food mechanism. This was accomplished by permitting each rat only 4 to 5 sec. of eating time. Most subjects quickly adapted to the sequence of pressing the lever, eating, and being removed.

[4] Amounts of urination and defecation have been regarded as measures of emotionality in many investigations, e.g., Hall (7).

## Results

### *Effects of Work and Inter-Trial Interval upon Response Times during Acquisition*

Figures 2 and 3 indicate the main trend of these results. In Fig. 2 the ordinate represents log mean response time, the abscissa successive blocks of 20 trials, the parameter work.[5] It is evident that the differential effect of work is immediate and lasting. The curves for the various work groups begin at different levels, never cross, and apparently approach different asymptotes. Figure 3 shows the effect of inter-trial interval upon acquisition. The differences are not as marked as those of the work groups; the most obvious is between the 20-sec. group and the other two. The differential effect of inter-trial interval increases as learning progresses. Thus, log mean response times increase with work and decrease with inter-trial interval.

[5] In this and all succeeding cases, "log" refers to common logarithms. Logarithms are used extensively in the present paper because the distributions of response times are positively skewed whereas those of log response times more nearly approximate normality.

The significance of the differences exhibited by these two sets of curves was tested by three analyses of variance. The first concerns performance during the initial ten trials. For each *S* the logarithm of the mean response time for the first ten trials was obtained. A summary of these scores is presented in Table I. An analysis of variance was performed on the transformed data. The only significant *F*-ratio is that for work conditions: it is 8.65, a value significant at beyond the .001 level for 2 and 36 *d.f.* During the early stages of learning, then, work exerted a marked influence on level of performance whereas inter-trial interval produced no differential effect.

An over-all test of the significance of the differences exhibited by the two sets of curves in Figs. 2 and 3 was afforded by a second analysis of variance. For each *S* the mean response time for each block of ten trials was obtained. These values were converted according to the formula,

$$X' = \log (X + 1).$$

Because some of the means are less than 1.00, this particular transformation was used to facilitate the analysis. The transformed scores were then averaged for each *S*. A summary of these data appears in Table II. An analysis of variance was performed on the set of obtained means. The *F* for work conditions is 14.96; that for inter-trial interval conditions, 7.93. Both values are significant at well be-

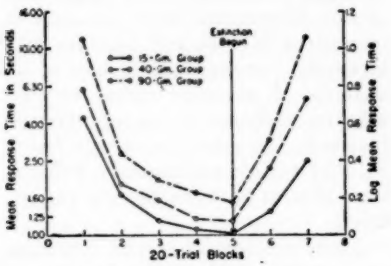


FIG. 2. Learning and extinction curves for the three work groups

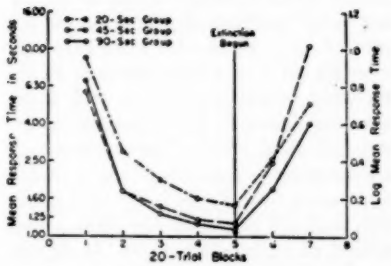


FIG. 3. Learning and extinction curves for the three inter-trial interval groups

TABLE I

AVERAGE LOG RESPONSE TIMES FOR THE VARIOUS EXPERIMENTAL GROUPS DURING THE FIRST TEN TRIALS OF ACQUISITION

| Work in Grams | Inter-Trial Interval in Seconds | | | |
|---|---|---|---|---|
| | 20 | 45 | 90 | Mean |
| 15 | 0.82 | 0.66 | 0.79 | 0.76 |
| 40 | 0.87 | 0.96 | 1.01 | 0.95 |
| 90 | 1.26 | 0.98 | 1.06 | 1.10 |
| Mean | 0.98 | 0.87 | 0.95 | 0.93 |

TABLE II

A SUMMARY OF THE DATA CONCERNING OVER-ALL PERFORMANCE DURING THE ACQUISITION SERIES *

| Work in Grams | Inter-Trial Interval in Seconds | | | |
|---|---|---|---|---|
| | 20 | 45 | 90 | Mean |
| 15 | 0.47 | 0.36 | 0.39 | 0.40 |
| 40 | 0.51 | 0.42 | 0.43 | 0.45 |
| 90 | 0.63 | 0.54 | 0.50 | 0.56 |
| Mean | 0.54 | 0.44 | 0.44 | 0.47 |

* These scores were obtained by computing for each *S* the mean response time for each block of ten trials, transforming these means according to the formula $X' = \log (X + 1)$, averaging the converted scores over the ten trial-blocks, and obtaining the mean of these means for each group.

yond the .01 level for 2 and 36 *d.f.* If these scores are taken as rough measures of level of performance during the entire course of acquisition, the results of this analysis indicate that performance during this period is favored by increased time between trials and by decreased bar-weight.

Inspection of Figs. 2 and 3 reveals that the learning curves for the various groups appear to approach different asymptotes. With the aid of the logarithmic transformation mentioned above, an analysis of variance was performed to test the significance of the differences among groups in level of performance during the last ten trials. Both work and inter-trial interval conditions produce significant effects: the *F* for the former, 4.17, is significant at the .05 level; that for the latter, 6.16, at the .01 level.

Because the trials were given in blocks of ten, with a rest period of either 15 min. (approximately) or 22 hr. between successive blocks, ample opportunity was provided for reminiscence to occur. In the present case, a *decrease* in response time on the first trial of any block, as compared to the response time on the last trial of the preceding block, would indicate that reminiscence occurred. However, the effect of reinforcement on trial 10 would also be a decrease in response time. In order to separate these two effects, i.e., to obtain a measure of reminiscence independent of the effect of the preceding reinforcement, the following analysis was made: The *gain* in response time of trial 10 over trial 9 for each block of 10 trials was obtained for each *S*, and these values averaged for the several work and inter-trial interval groups. These means represent the average effect of the reinforcement on trial 9 upon response time for trial 10. Similarly, the average gain in response time between the tenth trial of one block and the first of the next was computed. These means represent the average amount of reminiscence occurring between trial-blocks plus the average effect of reinforcement on trial 10. The former values were subtracted from the latter. This final set of scores represents amount of reminiscence independent of the preceding reinforcement, assuming that the reinforcements on trials 9 and 10 had roughly equivalent effects on response times. The values for the 15-, 40-, and 90-gm. groups are 0.06, −0.49, and −2.18; those for the 20-, 45-, and 90-sec. groups, −1.37, −1.06, and −0.17. Inspection of these means reveals that reminiscence did *not* occur. Instead, most groups exhibited an *increase* in response time after the rest period that presumably reflects a decrement in response strength. Although the magnitude of this decremental effect appears to increase with work and decrease with inter-trial interval, the results of an analysis of variance reveal no significant differences among groups. However, the mean decrement for all *S*s, 0.87 sec., is significantly different from zero at between the .02 and .01 level.

From the results of the preceding analyses it is evident that, in terms of response times during acquisition: (1) initial level of performance is an increasing function of work but is not related to inter-trial inteval; (2) over-all level of performance is an increasing function of work and a decreasing function of inter-trial interval; (3) final level of performance is an increasing function of work and a decreasing function of inter-trial interval; and (4) performance after a rest period of 15 min. or 22 hr. is significantly poorer (longer response time) than immediately before the rest pe-

riod when all groups are considered but does not differ among groups. Moreover, (5) because in each analysis of variance the interaction *F*-ratio is non-significant, it appears that the effects of work and inter-trial interval are independent of each other and additive, at least within the range of values employed (3).

*Effects of Work and of Inter-Trial Interval upon Response Times during Extinction*

Reference to Figs. 2 and 3 reveals, as might be expected, that response times during extinction follow a course opposite to that during acquisition. Increased work results in more rapid extinction; increased inter-trial interval, in slower extinction. The one exception occurs in the case of the 45-sec. group, which on the second day of extinction exhibits a log average response time greater than that of the 20-sec. group. Because of the extreme variability of the scores on the second day of extinction, it is not feasible to test the significance of the differences among groups for the total extinction period.[6] Table III contains a summary of the average log response times for the first 20 extinction trials. An analysis of variance of these data yields only one significant *F*-ratio: the *F* for work conditions is 6.47, a value significant at beyond the .01 level. If, as in the case of the acquisition data, these scores are taken as a rough measure of over-all level of performance during the first 20 extinction trials, the results indicate that increased work favors extinction but that variations in inter-trial interval produce no significant effect.

[6] There were indications in the behavior of the animals during extinction that frustration effects were present. Many rats became agitated, some defecated, etc. This may account, at least in part, for the increased variability of response times during extinction.

TABLE III

AVERAGE LOG RESPONSE TIMES FOR THE VARIOUS EXPERIMENTAL GROUPS DURING THE FIRST TWENTY EXTINCTION TRIALS

| Work in Grams | Inter-Trial Interval in Seconds | | | |
|---|---|---|---|---|
| | 20 | 45 | 90 | Mean |
| 15 | 0.53 | 0.38 | 0.30 | 0.40 |
| 40 | 0.61 | 0.49 | 0.60 | 0.57 |
| 90 | 0.80 | 0.87 | 0.49 | 0.72 |
| Mean | 0.65 | 0.58 | 0.46 | 0.56 |

The rest period of 22 hr. between the twentieth and twenty-first extinction trials provided an opportunity for spontaneous recovery to occur. If the response times on trial 21 are less than those on trial 20, then it can be said that spontaneous recovery occurred. A measure of amount of spontaneous recovery for the various groups was obtained in the following way: For each *S* the average response time for extinction trials 21 and 22 was subtracted from the average for trials 19 and 20. These values were then algebraically summed and averaged for the various experimental groups. A positive value indicates spontaneous recovery. The means for the 15-, 40-, and 90-gm. groups are 0.49, −2.91, and 2.18; those for the 20-, 45-, and 90-sec. groups are −1.02, 1.72, and −0.95. An analysis of variance yields no significant *F*-ratios. The average value for the entire group of 45 *S*s, −0.08, is not significantly different from zero. Thus there is no evidence for spontaneous recovery.

In brief, during extinction (1) work conditons produced differential effects on performance but inter-trial interval conditions did not, and (2) spontaneous recovery did not occur.

Analysis of variance is used exten-

sively in the present paper. It should be emphasized that inferences based upon the results of an analysis of variance are strictly valid only if it can be shown that the assumptions underlying this technique are fulfilled. The four assumptions that should be met are summarized concisely by Eisenhart (3). Examination of the data shows that in each of the analyses of variance reported in the present paper these assumptions are satisfactorily fulfilled. This was made possible by the transformations employed.

## Discussion

Hull (9, p. 300) states in his eighth postulate that the increment to reactive inhibition resulting from a single response evocation is an increasing function of the amount of work ($W$) involved in the response, and that reactive inhibition ($I_r$) dissipates as a decay function of time ($t'''$). When $W$ and $t'''$ are constant and $N$ trials are given in succession, it follows that the accumulation of $I_r$ is a negatively accelerated increasing function of $N$ provided that $t'''$ is not so great as to permit the complete dissipation of $I_r$ between responses. Furthermore, the rate of accumulation of $I_r$ is greater for larger amounts of $W$ and for smaller values of $t'''$.

In Postulate IX Hull (9, p. 300) asserts that conditioned inhibition ($_sI_r$) is generated as a secondary effect from $I_r$. However, Hull does not state clearly what conditions determine the amount of $_sI_r$ generated. Sometimes he seems to imply that the magnitude of the increment to $_sI_r$ resulting from a given response is directly proportional to the amount of $I_r$ present at the cessation of that response (9, p. 282); at other times he appears to regard the dissipation of $I_r$ to be the factor responsible for the production of $_sI_r$ (9, pp. 292–293, 298). Although these two statements are not incompatible, clarification is necessary if specific predictions are to be made.

According to Hull's fourth postulate (9, p. 178), whenever a stimulus and a response occur in close temporal contiguity and are closely associated in time with the diminution of a need, there results an increment to habit strength, i.e., the $S$—$R$ combination is reinforced; the magnitude of the reinforcement is an increasing function of the amount of need reduction. Now, if $I_r$ is a "primary negative drive," it must (in Hull's terms) result from a need that can be termed a "need for rest." Moreover, $I_r$ dissipation presumably reflects a diminution of this need. Therefore, stimuli present at the cessation of the response acquire a habit loading on $_sI_r$ the magnitude of which is a function of the amount of $I_r$ dissipating while $S$ is exposed to those stimuli. It was noted earlier that all $S$s were permitted to eat in the goal box for the same length of time, 5 sec., after pressing the lever. This means that the amount of $_sI_r$ generated after any given response was dependent upon the amount of $I_r$ that dissipated while $S$ was eating in the goal arm. Because the equation expressing the dissipation of $I_r$ as a function of time (9, p. 300) contains the decay constant $q$, it follows that the greater the amount of $I_r$ at the end of a response, the more will dissipate during the 5-sec. eating period, and so the greater will be the amount of $_sI_r$ generated. Thus, the faster reaction inhibition accumulates from trial to trial, the more rapidly conditioned inhibition builds up.

Hull states that $_sI_r$ is a habit and that the upper limit of habit strength is an increasing function of the mag-

nitude of reinforcement (9, p. 178). Therefore, under conditions in which $I_r$ accumulates from trial to trial it follows that amount of ${}_sI_r$ is a negatively accelerated increasing function of $N$ that rises more steeply as work is increased or inter-trial interval decreased.

There remains to be considered the effects of $I_r$ and ${}_sI_r$ upon response evocation. According to Postulate IX, reactive inhibition summates with conditioned inhibition to produce a net inhibitory potential ($\dot{I}_r$) that acts against reaction potential (${}_sI_r$) resulting in effective reaction potential (${}_s\bar{E}_r$) (9, p. 300). These relations are described by the following equations:

$$\dot{I}_r = I_r + {}_sI_r$$
$${}_s\bar{E}_r = {}_sE_r - \dot{I}_r$$

Response strength as measured is a direct function of ${}_s\bar{E}_r$. For the present experimental situation these equations imply that, because all variables contributing to the development of ${}_sE_r$ (e.g., drive, number of reinforcements, amount of reinforcement, delay in reinforcement) are equated, ${}_s\bar{E}_r$ depends upon $\dot{I}_r$. Effective reaction potential, which is responsible for response evocation, is thus an increasing function of inter-trial interval and a decreasing function of work.

It is now possible to make several predictions concerning the effects of work and inter-trial interval upon performance during acquisition in the present experiment. Because $I_r$ should develop differentially for the various experimental groups, being larger for those doing greater amounts of work at shorter inter-trial intervals, it is predicted that at any given time during the acquisition series (1) level of performance will be a decreasing function of work and (2) an increasing function of inter-trial interval. Thus, response time should increase with bar-weight and decrease with inter-trial interval. The results confirm these predictions (Figs. 2 and 3) with but one exception: inter-trial interval conditions did not produce significant differential effects during the early stages of learning. This discrepancy will be discussed later.

Because the learning series consisted of successive blocks of ten trials each, it is expected that during any trial-block both $I_r$ and ${}_sI_r$ will accumulate, and that between trial-blocks $I_r$ will dissipate. Thus on the initial trial of any given block only conditioned inhibition will be present and effective reaction potential will be increased by an amount proportional to the amount of reactive inhibition that has dissipated. On this basis it is expected that between the last trial of one block and the first trial of the next there will occur a decrease in response time over and above that attributable to reinforcement—the phenomenon of "reminiscence." The amount of reactive inhibition that accumulates during a trial-block and dissipates in the ensuing rest period will increase with bar-weight and decrease with inter-trial interval. Therefore, amount of reminiscence is predicted to be (3) an increasing function of bar-weight and (4) a decreasing function of time between trials. Neither of these predictions is confirmed. In fact, instead of a decrease in response time after the rest period there occurred on the average a net increase, i.e., a decrement in response strength, the amount of which does not differ significantly among groups.

This finding can be explained in the following way: In the present experiment it is probable that the after-effects of any given trial, e.g., taste traces and proprioceptive traces, form part of the stimulus situation on the next trial; hence, the bar-pressing response is conditioned to such traces. Data showing that proprioceptive conditioning can occur have been reported by Grice (6). On the initial one or two trials of any block these stimuli are largely absent, and, according to the principle of stimulus generalization (9, p. 199), habit strength is weakened. If the resulting

sively in the present paper. It should be emphasized that inferences based upon the results of an analysis of variance are strictly valid only if it can be shown that the assumptions underlying this technique are fulfilled. The four assumptions that should be met are summarized concisely by Eisenhart (3). Examination of the data shows that in each of the analyses of variance reported in the present paper these assumptions are satisfactorily fulfilled. This was made possible by the transformations employed.

## Discussion

Hull (9, p. 300) states in his eighth postulate that the increment to reactive inhibition resulting from a single response evocation is an increasing function of the amount of work ($W$) involved in the response, and that reactive inhibition ($I_r$) dissipates as a decay function of time ($t'''$). When $W$ and $t'''$ are constant and $N$ trials are given in succession, it follows that the accumulation of $I_r$ is a negatively accelerated increasing function of $N$ provided that $t'''$ is not so great as to permit the complete dissipation of $I_r$ between responses. Furthermore, the rate of accumulation of $I_r$ is greater for larger amounts of $W$ and for smaller values of $t'''$.

In Postulate IX Hull (9, p. 300) asserts that conditioned inhibition ($_sI_r$) is generated as a secondary effect from $I_r$. However, Hull does not state clearly what conditions determine the amount of $_sI_r$ generated. Sometimes he seems to imply that the magnitude of the increment to $_sI_r$ resulting from a given response is directly proportional to the amount of $I_r$ present at the cessation of that response (9, p. 282); at other times he appears to regard the dissipation of $I_r$ to be the factor responsible for the production of $_sI_r$ (9, pp. 292–293, 298). Although these two statements are not incompatible, clarification is necessary if specific predictions are to be made.

According to Hull's fourth postulate (9, p. 178), whenever a stimulus and a response occur in close temporal contiguity and are closely associated in time with the diminution of a need, there results an increment to habit strength, i.e., the $S$—$R$ combination is reinforced; the magnitude of the reinforcement is an increasing function of the amount of need reduction. Now, if $I_r$ is a "primary negative drive," it must (in Hull's terms) result from a need that can be termed a "need for rest." Moreover, $I_r$ dissipation presumably reflects a diminution of this need. Therefore, stimuli present at the cessation of the response acquire a habit loading on $_sI_r$ the magnitude of which is a function of the amount of $I_r$ dissipating while $S$ is exposed to those stimuli. It was noted earlier that all Ss were permitted to eat in the goal box for the same length of time, 5 sec., after pressing the lever. This means that the amount of $_sI_r$ generated after any given response was dependent upon the amount of $I_r$ that dissipated while $S$ was eating in the goal arm. Because the equation expressing the dissipation of $I_r$ as a function of time (9, p. 300) contains the decay constant $q$, it follows that the greater the amount of $I_r$ at the end of a response, the more will dissipate during the 5-sec. eating period, and so the greater will be the amount of $_sI_r$ generated. Thus, the faster reaction inhibition accumulates from trial to trial, the more rapidly conditioned inhibition builds up.

Hull states that $_sI_r$ is a habit and that the upper limit of habit strength is an increasing function of the mag-

nitude of reinforcement (9, p. 178). Therefore, under conditions in which $I_r$ accumulates from trial to trial it follows that amount of ${}_sI_r$ is a negatively accelerated increasing function of $N$ that rises more steeply as work is increased or inter-trial interval decreased.

There remains to be considered the effects of $I_r$ and ${}_sI_r$ upon response evocation. According to Postulate IX, reactive inhibition summates with conditioned inhibition to produce a net inhibitory potential ($\dot{I}_r$) that acts against reaction potential (${}_sI_r$) resulting in effective reaction potential (${}_s\bar{E}_r$) (9, p. 300). These relations are described by the following equations:

$$\dot{I}_r = I_r + {}_sI_r$$
$${}_s\bar{E}_r = {}_sE_r - \dot{I}_r$$

Response strength as measured is a direct function of ${}_s\bar{E}_r$. For the present experimental situation these equations imply that, because all variables contributing to the development of ${}_sE_r$ (e.g., drive, number of reinforcements, amount of reinforcement, delay in reinforcement) are equated, ${}_s\bar{E}_r$ depends upon $\dot{I}_r$. Effective reaction potential, which is responsible for response evocation, is thus an increasing function of inter-trial interval and a decreasing function of work.

It is now possible to make several predictions concerning the effects of work and inter-trial interval upon performance during acquisition in the present experiment. Because $\dot{I}_r$ should develop differentially for the various experimental groups, being larger for those doing greater amounts of work at shorter inter-trial intervals, it is predicted that at any given time during the acquisition series (1) level of performance will be a decreasing function of work and (2) an increasing function of inter-trial interval. Thus, response time should increase with bar-weight and decrease with inter-trial interval. The results confirm these predictions (Figs. 2 and 3) with but one exception: inter-trial interval conditions did not produce significant differential effects during the early stages of learning. This discrepancy will be discussed later.

Because the learning series consisted of successive blocks of ten trials each, it is expected that during any trial-block both $I_r$ and ${}_sI_r$ will accumulate, and that between trial-blocks $I_r$ will dissipate. Thus on the initial trial of any given block only conditioned inhibition will be present and effective reaction potential will be increased by an amount proportional to the amount of reactive inhibition that has dissipated. On this basis it is expected that between the last trial of one block and the first trial of the next there will occur a decrease in response time over and above that attributable to reinforcement—the phenomenon of "reminiscence." The amount of reactive inhibition that accumulates during a trial-block and dissipates in the ensuing rest period will increase with bar-weight and decrease with inter-trial interval. Therefore, amount of reminiscence is predicted to be (3) an increasing function of bar-weight and (4) a decreasing function of time between trials. Neither of these predictions is confirmed. In fact, instead of a decrease in response time after the rest period there occurred on the average a net increase, i.e., a decrement in response strength, the amount of which does not differ significantly among groups.

This finding can be explained in the following way: In the present experiment it is probable that the after-effects of any given trial, e.g., taste traces and proprioceptive traces, form part of the stimulus situation on the next trial; hence, the bar-pressing response is conditioned to such traces. Data showing that proprioceptive conditioning can occur have been reported by Grice (6). On the initial one or two trials of any block these stimuli are largely absent, and, according to the principle of stimulus generalization (9, p. 199), habit strength is weakened. If the resulting

decrement in response strength was greater than the increment produced by the dissipation of reactive inhibition, then the obtained increase in response times is explained.

The fact that the interaction *F*-ratio is non-significant in each analysis of variance indicates that the effects of work and inter-trial interval upon performance during acquisition (as measured by the transformed scores) are independent and additive. Kimble and Bilodeau (11) have reported a similar finding for motor learning in human subjects.

Extinction, according to Hull, is explained in terms of the accumulation of both reactive and conditioned inhibition (9, p. 298). When their total inhibitory power equals or exceeds the magnitude of reaction potential, no response occurs. If a rest period is given, spontaneous recovery occurs: $I_r$ dissipates, and effective reaction potential regains part of its former strength. The present experiment is designed to test several predictions about the process of extinction. When reinforcement is withheld, habit strength—hence, reaction potential—remains constant whereas net inhibitory potential continues to increase in the same way as during acquisition. Therefore, at any given time during extinction level of performance, as measured by response time, should be (5) an increasing function of work and (6) a decreasing function of inter-trial interval.[7] The results show that the first of these predications is verified but that the second is not. The former finding confirms results reported by Fitts (4), James (10), and Mowrer and Jones (13); the latter, which will be discussed later, agrees with results reported by several investigators, e.g., Gagné (5) and Porter (14), who have studied the extinction of simple operant responses under conditions of varied inter-trial interval.

The 22-hr. rest period given between extinction trials 20 and 21 provided an opportunity for spontaneous recovery to occur. According to Hull, this phenomenon is based on the same mechanism as is reminiscence: reactive inhibition dissipates during a rest period, and response strength is proportionally increased. As was shown earlier, the amount of reactive inhibition that accumulates during a trial-block increases with work and decreases with inter-trial interval. Therefore, amount of spontaneous recovery should be (7) an increasing function of bar-weight and (8) a decreasing function of time between trials. Neither prediction is confirmed by the results: the rest period produced no changes in response times. These negative findings can be accounted for on the same basis as those concerned with reminiscence, i.e., in terms of the weakening of the response owing to a change in the stimulus situation during the initial few trials after the rest period. The fact that there was no decrement in response strength after the rest period in extinction whereas there was such a decrement during acquisition is explained by the greater number of trials in the trial-blocks during extinction, i.e., more reactive inhibition accumulates during 20 trials than during 10, and its dissipation results in a greater strengthening of the response after a rest period.

The results of the present study show that during the early stages of both acquisition and extinction the differential effects of inter-trial interval upon performance are negligible whereas those of work are marked (Figs. 2 and 3). It is possible that the range of amounts of inhibition produced by the work variable is much greater than that produced by the time variable. If this is the case, then a greater number of trials would be needed to detect the influence of the latter variable, i.e., the differential effects of inter-

[7] The objection might be made that the present experimental design does not permit an adequate test of the effects of work and inter-trial interval upon performance during extinction because the groups were trained under different conditions of work and inter-trial interval. However, according to Hull's theory all groups should have (1) developed equal habit strengths during acquisition and (2) differed at the beginning of extinction only in amount of conditioned inhibition. Hence, because extinction is said to occur because of the continued development of inhibition, it is argued that the design is adequate.

trial interval would increase with number of trials. That this trend did occur may be seen in Fig. 3.

The extinction results pose a more difficult problem. The preceding explanation is not applicable unless the initial values of habit strength and conditioned inhibition are roughly equal for the several groups. However, analysis of the data reveals that there was a strong tendency toward equalization of response strengths during the initial extinction trials. An analysis of variance of the mean response time scores (transformed as before) on the first five trials of extinction yields no significant *F*-ratios. It is probable that the change in the stimulus situation produced by the removal of reinforcement resulted in a weakening of both habit strength and conditioned inhibition, thus bringing about an equalization of response strengths among groups.[8] Further, because the stimulus situation was permanently changed, a *new* ${}_{s}I_{r}$ habit arose at that new point on the stimulus generalization continuum. If this is the case, and if the range of inhibition values produced by bar-weight is greater than that produced by inter-trial interval, then, just as in acquisition, a greater number of trials would be required to detect differential effects of inter-trial interval upon performance during extinction.

Fortunately, data affording a test of this prediction are available for 25 rats that completed 60 extinction trials. The average response times of the 20- and 90-sec. groups for trials 1–10 of extinction are 1.25 and 1.49, values not significantly different; the corresponding means for trials 51–60 are 6.52 and 3.92 sec., values significantly different at beyond the .01 level. These data suggest that the differential effects of inter-trial interval upon performance do increase as extinction proceeds. Thus, it appears probable that the negligible effect of inter-trial interval conditions upon performance during the early stages of both acquisition and extinction is related to the range of values of inhibition produced by these conditions.

It should be pointed out that the results of the present investigation tend to substantiate the findings of the only other comparable experiment, that of Kimble and Bilodeau (11). The specific points of agreement are the following: (1) amount of work is a more significant factor than length of inter-trial interval with respect to level of performance during learning; (2) the initial differential effects of inter-trial interval upon level of performance are small, becoming increasingly larger as acquisition proceeds; and (3) the effects of work and of inter-trial interval upon performance during acquisition are independent and additive. Because the Kimble-Bilodeau study utilized human instead of animal subjects, an entirely different experimental situation, and a different measure of performance, the fact that similar results emerged from the two experiments strongly suggests that the same mechanisms are operative in human and animal learning and that similar explanations apply.

## Summary

The most comprehensive explanation of the effect of different amounts of effort, viewed either as absolute amount of work per response or as temporal distribution of work, upon the processes of acquisition and extinction is that offered by Hull (9). The two postulates in Hull's system most pertinent to this problem are numbers VIII and IX, in which the

[8] Because conditioned inhibition is presumed to be a negative habit of "not-responding," it is plausible to assume that it is a special case of the more general class of avoidance behavior. It has been shown that the spatial gradient of avoidance is steeper than that of approach (1). If a similar assumption is made here, i.e., that the generalization gradient of ${}_{s}I_{r}$ is steeper than that of ${}_{s}H_{r}$, then the explanation of the tendency toward equalization of response strengths is greatly strengthened.

concepts of "reactive inhibition" and "conditioned inhibition" are set forth. The present study attempts to provide a more crucial test of these postulates: eight predictions derived from them are tested in an experiment in which both of the specified independent variables, amount of work and inter-trial interval, are manipulated *simultaneously* and *independently of each other* in the same situation.

A total of 45 male albino rats was run in a simple maze-running–lever-pressing situation. Five *S*s were assigned at random to each of nine experimental groups formed by casting three values of work and three of inter-trial interval into a simple factorial design. Changes in response strength during 100 acquisition trials, given in blocks of 10, and 40 extinction trials, given in blocks of 20, were measured in terms of latency of the bar-pressing response. The principle method used to analyze the data is analysis of variance. The main results are the following:

1. During the first ten acquisition trials response latency is an increasing function of work but is not related to inter-trial interval.

2. During the course of acquisition response latency is an increasing function of work and a decreasing function of inter-trial interval.

3. During the last ten acquisition trials response latency is an increasing function of work and a decreasing function of inter-trial interval.

4. During acquisition the effects of work and of inter-trial interval upon response latency are independent and additive.

5. During acquisition the rest period after each block of ten trials did not produce reminiscence; instead, it produced a decrement in response strength the magnitude of which did not differ among groups.

6. During the first 20 extinction trials response latency is an increasing function of work but is not related to inter-trial interval.

7. As in acquisition, the differential effects of work upon response latency during extinction are greater than those of inter-trial interval, the latter becoming progressively larger as the number of trial is increased.

8. During extinction the rest period after the first block of 20 trials did not produce spontaneous recovery.

9. The experimental results do not seem to be related to a lack of adequate control of such variables as motivation, body weight, emotionality, activity level, pre-training procedure, differential environmental cues, or extraneous stimulation.

Three of the eight predictions are fully confirmed by the results. Theoretical explanations, consistent with Hull's theory, are offered to account for the negative findings.

(Manuscript received
December 20, 1949)

## References

1. Brown, J. S. Gradients of approach and avoidance responses and their relation to level of motivation. *J. comp. physiol. Psychol.*, 1948, 41, 450–465.
2. Denny, R. M. A demonstration of learning based solely on the work inhibition drive —a substantiation of Hull's concept of *Ir*. *Amer. Psychologist*, 1948, 3, 250–251.
3. Eisenhart, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, 3, 1–21.
4. Fitts, P. M. Perseveration of non-rewarded behavior in relation to food-deprivation and work-requirement. *J. genet. Psychol.*, 1940, 57, 165–191.
5. Gagné, R. M. The effect of spacing of trials on the acquisition and extinction of a conditioned operant response. *J. exp. Psychol.*, 1941, 29, 201–216.
6. Grice, G. R. The relation of secondary reinforcement to a delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, 38, 1–16.

7. HALL, C. S. Drive and emotionality; factors associated with adjustment in the rat. *J. comp. Psychol.*, 1934, 17, 89–108.
8. HILGARD, E. R., AND MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century Co., 1940.
9. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century Co., 1943.
10. JAMES, W. T. Experimental observations indicating the significance of work on conditioned motor reactions. *J. comp. Psychol.*, 1941, 32, 353–366.
11. KIMBLE, G. A., AND BILODEAU, E. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, 39, 150–157.
12. MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green and Co., 1946.
13. MOWRER, O. H., AND JONES, H. M. Extinction and behavior variability as function of effortfulness of task. *J. exp. Psychol.*, 1943, 33, 369–386.
14. PORTER, J. M., JR. Extinction of an acquired response as a function of the interval between successive non-rewarded runs. *J. comp. Psychol.*, 1938, 26, 261–270.
15. ROHRER, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, 37, 473–493.
16. SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, 55, 1–40.
17. THOMPSON, M. E. Learning as a function of the absolute and relative amounts of work. *J. exp. Psychol.*, 1944, 34, 506–515.

# BIDIRECTIONAL GRADIENTS IN THE STRENGTH OF A GENERALIZED VOLUNTARY RESPONSE TO STIMULI ON A VISUAL–SPATIAL DIMENSION

BY JUDSON S. BROWN, EDWARD A. BILODEAU, AND MARTIN R. BARON[1]

*State University of Iowa*

One of the common observations made in conditioning experiments is that the conditioned response is elicitable not only by the original conditioned stimulus but also by other (nonconditioned) stimuli. In general, the greater the disparity between the conditioned stimulus and the test stimulus, the weaker or less frequent will be the generalized response to the test stimulus. An examination of the studies in which this phenomenon has been systematically explored reveals that two rather different methods have been used in varying the degree of disparity between the two stimuli. The first and most direct method consists simply in changing certain of the physically defined properties of the stimulus. Thus, for example, Hovland (5, 6) conditioned the galvanic skin response to a tone and then tested for generalization effects by presenting other tones whose frequency or intensity differed from that of the original tone. Similarly, Brown (2), in studying the generalization of approach responses in rats, modified the intensity of the visual stimulus toward which the animals had been trained to run.

The second method of varying the disparity between the conditioned and nonconditioned stimuli has been employed primarily in studies where the stimuli were tactile in character. It differs from the first method in that changes are not introduced in the physical properties of the stimulus (usually a vibrator) but only in the point of its application to the skin. Bass and Hull's study of generalization (1) is typical. They conditioned the galvanic skin response in human subjects to a vibratory tactile stimulus and tested for degree of generalization by applying identical vibrators to the skin at increasingly remote positions from the point originally conditioned. In spite of the physical identity of the vibrators, gradients in the amplitude of the galvanic response were nevertheless obtained, response amplitude decreasing with distance. In a somewhat comparable study, Gibson (3) observed that analogous generalization-like phenomena could be obtained with voluntary verbal responses to tactile stimuli. It would appear, therefore, that shifting the spatial position of a stimulus on the skin changes the effect of that stimulus in a manner that is functionally equivalent to a direct change in its physical properties. Studies employing this second method may be described as experiments on *spatial generalization* since the spatial separation of the test stimuli on the skin becomes, in effect, the independent variable against which variations in strength of response are plotted.

Inasmuch as the major studies of spatial generalization have used tactile stimuli, the question arises as to whether the phenomenon is restricted to such stimuli, or whether comparable results might be obtained with

[1] Second author now at the Human Resources Research Center, Lackland Air Force Base, Texas. Third author now at Kent State University.

visual stimuli whose position in space is systematically varied. The experiments reported herein were designed to provide at least a partial answer to this question. The following general procedure was employed. The subjects were seated in a semidarkened room in front of a horizontal row of seven lamps. They were instructed to react as quickly as possible to the lighting of the central lamp by lifting their fingers from a reaction key. They were told not to react to the lighting of other lamps. Following an initial series of trials in which only the central lamp was turned on, a test series was given in which each of the other six lamps was occasionally turned on. It was assumed that the initial series of trials would build up a strong tendency to react to the central light and that this tendency would generalize or spread to adjacent lights in proportion to their spatial nearness to the central lamp.

## Apparatus

The major component of the apparatus was the stimulus-lamp board, a thin plywood panel 6 ft. long and 30 in. wide, mounted on its long edge on a standard laboratory table. Seven small, frosted lamps (115 v., 7.5 w.) were fastened to the panel in a horizontal row 4 ft. from the floor. The lamps were uniformly spaced at 8° intervals. The panel was curved so that all lamps would be equidistant (5 ft.) from the eyes of the subjects when they were seated directly in front of the central lamp. The height of the subject's chair could be adjusted so that the row of lamps would be approximately at eye level. A green-jeweled, 6.3-v. radio pilot lamp, located 2 in. directly above the central lamp, served as both a ready signal and as a visual fixation point. The reaction key was mounted on a stand at a convenient height (30 in.) in front of the subject. By the use of suitable switching arrangements, *E*, who was seated behind the stimulus panel, could turn on any one of the seven lamps following the lighting of the green ready signal. The frequency of responses to all lights was tabulated and their latency was measured to the nearest one-hundredth second by means of a Standard Electric Timer.

The apparatus was situated in a windowless, black-walled room. The major source of illumination was that provided by a 7.5-w. lamp located 6 ft. above, and directly over, the central stimulus lamp. This provided sufficient light so that *S*s could see the stimulus panel and the lamps but little of such details as lamp sockets and connecting wires, which, like the panel, were painted a flat black. Some slight additional illumination was provided by another small, shaded work lamp located behind the panel.

## Procedure

Two experiments involving almost identical procedures were carried out with this apparatus. The first was conducted as a class experiment in which 24 *S*s were tested by 24 different *E*s. The second was carried out in a more conventional manner, with 24 additional *S*s being tested by a single *E*. Since the procedure of the first experiment was necessarily somewhat more variable than that of the second, only the latter will be described in detail.

The *S*s of the second experiment were 24 undergraduate and graduate students in psychology, half of whom were men and half women. After *S*s were correctly seated, they were told that the experiment was one on reaction time and their task was to react as quickly as possible to the lighting of the central lamp by lifting their fingers momentarily from the reaction key. They were told that other lamps would occasionally be turned on to which they were not to respond and that if they should react falsely on occasion they were not to be unduly concerned but were to keep on reacting as quickly and accurately as possible.

Following the reading of the instructions, *S*s were given an initial series of 35 training trials with only the central lamp being lighted. Successive trials in this series were given at intervals of between 10 and 15 sec. The foreperiod between the onset of the ready signal and the lighting of the central light was varied in a predetermined random order from 3 to 5 sec. The training trials were followed without interruption or warning by a test series in which each of the three lights to the left and right of the central lamp was presented four times. These 24 test trials were interspersed among 104 additional trials in reacting to the central light. At least three but never more than seven trials with the central light intervened between successive test trials with the peripheral lights. Six different orders were used for the test trials, each order beginning with the presentation of a different one of the six test lights. Four subjects were tested under each of these six orders.

TABLE I

RESPONSES TO LIGHTS IN EXPERIMENTS 1 AND 2

The first two rows give the percentages of responses made by all subjects to the central and peripheral lights in the first and second experiments, respectively. The third row gives the percentage of subjects in both experiments who reacted at least once to the lights.

| Percentages | Visual Angle between Central and Peripheral Lights | | | | | | |
|---|---|---|---|---|---|---|---|
| | −24° | −16° | −8° | 0 | +8° | +16° | +24° |
| % R's | | | | | | | |
| Exper. 1 | 34.4 | 35.4 | 42.7 | 100 | 42.7 | 34.4 | 28.2 |
| Exper. 2 | 27.1 | 38.5 | 46.9 | 100 | 44.8 | 33.4 | 27.1 |
| % Ss Responding | | | | | | | |
| Exper. 1 and 2 | 66.8 | 77.1 | 81.5 | 100 | 81.5 | 77.1 | 60.5 |

The procedure of the first (class) experiment differed from that of the second in that 50 training trials were used instead of 35, a pre-training session was given in which responses to all lights were required, somewhat fewer trials to the central light were interspersed between test trials, and all 24 subjects were undergraduate students.

## RESULTS

The principal results of the two experiments as revealed by the percentage of reactions made to the test and central lights are presented in the first two rows of the table and in the figure. From an inspection of these data, it is evident that frequency of response decreases progressively on each side of the central light as the angle separating a given test stimulus from the center is increased. Presumably, the instructions to respond only to the central light, coupled with the original and interspersed training trials, resulted in a strong tendency to react to that light. A generalization-like spread of the tendency is then exhibited, those lights near the center of the visual field being affected more than those near the periphery. This effect, it should be noted, is observed in both experiments and on both sides of the central position, in spite of the fact that the instructions prohibited responses to *all* peripheral lights. Had these instructions been completely effective, no responses would have been made to any of the peripheral lights. The occurrence of numerous "false" responses suggests, therefore, the presence of a strong set to respond to any light signal, and the fact of a gradient *among the prohibited lights*[2] indicates that those stimuli nearest the one on which extensive training had been

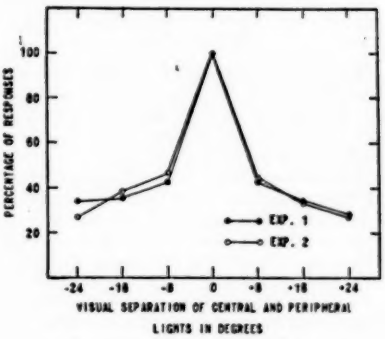


FIG. 1. Responses to lights in Experiments 1 and 2

[2] The initial drops in the gradients, from the central light to the 8° points on each side, can be interpreted as due both to the differential instructions at the points compared and to postulated decreases in the strengths of the generalized reaction tendencies. The further drops from the 8° to the 16° points, however, cannot be attributed to the instructions, since these applied equally to *all* peripheral lights.

given are more affected by the generalized positive reaction tendency than are more remote lights.

It may be observed from the third row of the table that a symmetrical bidirectional gradient is also obtained by computing the percentage of subjects (both experiments combined) who reacted at least once to the stimulus lights. This may be taken as further support for a gradient of response strength, since it indicates that the gradients obtained by calculating percentages of responses were not due solely to the reactions of a few atypical subjects. It can also be shown that if only those subjects who reacted at least once to each light are treated as a separate group, their mean response frequencies also exhibit a gradient (with one slight inversion) on each side of the central point.

It may be tentatively concluded, therefore, that the results of the present studies, expressed in terms of frequency of response,[3] indicate that if training is given in reacting to a stimulus at the center of the visual field, generalized responses will occur to peripheral stimuli, the tendency to react to such stimuli decreasing as they become more peripheral. Since this effect was observed in two separate experiments and on both sides of the central stimulus lamp, no statistical proof of the fact of a gradient is deemed necessary.[4]

## Discussion [5]

It has been shown in the present paper, on the basis of two separate experiments, that a phenomenon paralleling the generalization of conditioned responses to tactual stimuli can be obtained with visual stimuli whose positions in space are varied. This spatial generalization effect was observed with a "voluntary" response of lifting the fingers from a reaction key when responses to all stimuli save the middle one of seven were prohibited by instructions. It would appear, therefore, that visual space constitutes a suitable dimension within which further, more systematic, studies of generalization might well be pursued with profit.

Although the majority of previous studies of generalization have followed classical conditioning methods with nonvoluntary responses, there appears to be little reason for sup-

[3] Although reaction times were also recorded for each response, a detailed analysis of the data showed no consistent trends. The distributions of individual mean scores at each light were extremely skewed and irregular, one or two being very nearly bimodal. Because of the marked instability of these data, it seems unwise to speculate regarding the probable nature of the generalization curve as revealed by the response latency measure. Such hints as do emerge from the analysis appear to favor the notion that reaction times are longest for the central light and decrease with increasing distance from that position. A result of this sort, which is supported to some degree by Gibson's (3) latency data, could well be due to the combined effect of the instructions (which permit long latency reactions to the "approved" stimulus) and to the possibility (as suggested by Gibson) that the subjects are capable of inhibiting potential long-latency responses to peripheral stimuli.

[4] Further confirmation of these results was obtained in an additional class experiment performed since the preparation of this paper. The third experiment paralleled the others in all major respects save that the subjects were required to push down on a switch lever at the appearance of the center light and to lift up the lever at the appearance of any other light. The percentages of pressing responses made by the subjects ($N = 24$) to each of the seven lights from left to right were, 31.2, 36.4, 47.9, 100, 44.8, 33.4, and 30.2, respectively. A comparison of these values with those listed in the accompanying table reveals a surprising degree of consistency.

[5] The authors are indebted to Drs. I. E. Farber and K. W. Spence for helpfully critical comments on much of the material in this section.

posing that the basic phenomenon of generalization is unique to these procedures. As has been noted above, functionally similar relationships have been obtained with verbal responses (3) and (using animals) with instrumental conditioning techniques (2). Behavior such as lifting the fingers from a key, when elicited by instructions to adult subjects, is certainly susceptible to the influence of complex attitudes and sets, but it is questionable whether it is a phenomenon of an entirely different order from that of a well-learned conditioned response. Thus, our instructions to respond only to the central light could be closely paralleled within classical conditioning by the use of preliminary *differential* training during which presentations of the central light, but not the peripheral lights, would be followed by reinforcement. Additional conditioning trials could then be given to the central light (corresponding to the *initial* training trials of the present studies) and followed by tests for generalized responses to the (equally extinguished) peripheral lights. If gradients in the strength or frequency of responses to the nonreinforced lights were obtained, the parallel would be complete.

Since the gradients of response frequency reported herein have been termed "generalization-like" gradients, it is desirable to consider the question of whether this use of the term generalization is justified or whether the observed phenomenon is something entirely different.

In treating this problem, it is necessary to insist at the outset that a distinction must be made between the *empirical phenomenon of generalization* on the one hand and *theories or hypotheses about generalization on the other.* The empirical phenomenon of stimulus generalization is simply this: an organism that has been trained (or instructed) to respond to a designated stimulus will also respond, under certain specifiable conditions, to formerly neutral stimuli on which no training has been given (or to stimuli to which responses have been prohibited by instructions). This, it is believed, is the fundamental meaning of the term as derived from numerous experiments in classical, instrumental, and "voluntary" conditioning. It serves, in effect, as a simple operational definition of empirical generalization. A *theory* of empirical generalization, however, is something quite different; it consists of guesses or hypotheses as to the factors or processes underlying the observable data.[6] In recent years,

[6] It should be noted that neither Hull nor Spence has ever seriously attempted to develop a neurophysiological theory of the mechanisms underlying empirical generalization. In an early paper on the subject, Hull specifically states that primary generalization is to have the status of a postulate with no assumptions being made as to its origin (8, p. 12). Although it is true that in his *Principles of Behavior* he considers briefly the possibility of adopting Guthrie's theory (4), which involves afferent activities common to both the conditioned and nonconditioned stimuli, he concludes that the hypothesis is not readily testable and that ". . . we must get along as best we can with a molar analysis based on empirically determined functional relations" (7, p. 191). Hull has proceeded by simply postulating, on the basis of empirical generalization data, that training to respond to one stimulus leads to the development of associative connections (habit strength) for other stimuli in addition to the training stimulus. Within his system, the making of such an assumption is an inescapable logical necessity, not an ad hoc postulate specific to generalization. Since he assumes that any response is a function of exictatory potentiality (${}_sE_r$), and that ${}_sE_r$ is in turn uniquely determined by the multiplicative combination of habit strength (${}_sH_r$) and drive ($D$), it follows that whenever a response *does* occur to a nonconditioned stimulus some habit strength must be posited to maintain a consistent system. From this it would appear that insofar as Hull can be said to have a specific theory of generalization, it consists of his equation for the curve of generalized habit strength,

extensive misunderstanding and controversy have resulted from a failure to make this and other terminological distinctions clear. In some of the current polemics against "generalization" (9, 10, 11), for example, the terms *generalization*, *stimulus generalization*, and *primary stimulus generalization* have been used quite indiscriminately to mean one or more of the following: (*a*) the *empirical phenomenon of generalization* as defined above, (*b*) a *gradient in the strength or frequency of responses* to nonconditioned stimuli, (*c*) the *transfer of differential response proficiency* from one discrimination problem involving both positive and negative stimuli to a second problem of the same sort, (*d*) an *irradiation or spread of neural excitation* within the cortex, (*e*) a *postulated growth of associative strength* between a conditioned response and a group of nonconditioned stimuli, (*f*) a *gradient in the strength of the postulated associative tendencies*. Unfortunately, it is often difficult, if not impossible, to determine precisely which of these meanings is intended in any specific instance, even by reference to the surrounding contextual material. As a consequence, arguments directed apparently against the heuristic value of postulates such as (*e*) or (*f*), or against a physiological theory of empirical generalization such as (*d*), suffer marked losses in intelligibility and precision.[7]

If the present studies are considered in the light of these distinctions it is evident that the experimental findings may be properly regarded as instances of *empirical generalization*. The subjects were given practice in responding to the middle light and were told not to respond to the peripheral lights. Nevertheless, when the peripheral lights were presented, the subjects responded to them much as if they were the central light. Whether these responses were due to "cortical irradiation," to "errors of perception," or to the "spread of habit strength" has no bearing whatever upon the empirical facts as such. And the proof or disproof of any or all of these hypotheses is likewise irrelevant with respect to the validity of the experimental data.

Since the present authors would probably follow Hull (7) in *postulating* that gradients of associative strength underlie empirical generalization gradients, it is important to examine briefly some of the arguments that have been advanced by Lashley and Wade (9) and by Razran (10, 11) against the utility of such an assumption. The essence of one of these contentions is that much, if not all, of empirical stimulus generalization is due, not to the growth of associative strengths to nonconditioned stimuli but to either (*a*) a failure to note distinguishing characteristics of the stimulus, or to (*b*) a failure to associate such distinguishing characteristics with the conditioned reaction (9, p. 81; 10, p. 359). Let us consider how these two statements might apply to the present study.

To say that generalization is attributable to a *failure to note distin-*

together with all of the other aspects of his system which specify the nature of habit strength, drive, and excitatory potential, and their relations to antecedent conditons and behavior. Spence has also simply postulated that positive and negative excitatory tendencies do generalize (12, p. 433), and has shown that if additional specific assumptions are made as to the form of the gradients, it then becomes possible to deduce certain phenomena of discrimination learning and transposition.

[7] There is also considerable room for improvement in the terminological usage of writers who support the concept of "generalization," particularly with regard to the making of a clear distinction between empirical and associative (habit-strength) generalization (2, 3, 7, 12).

*guishing characteristics of the stimulus* is inescapably circular, as it stands. A reliable inference that a subject has failed to note a particular stimulus characteristic can only be made if the subject fails to respond differentially to that characteristic. But a failure to respond differentially is also the central fact of (complete) empirical generalization; the subject reacts to the nonconditioned stimulus as if it were the conditioned stimulus. The "closure" is thus complete; an observed failure to respond differentially (i.e., a failure to note) is proposed as the explanation of an observed failure to respond differentially (i.e., the making of a generalized response). It is evident that if one is to use failure-to-note to explain empirical generalization, the failure-to-note must be defined independently of the particular generalization reaction it is supposed to explain. Such an independent definition has apparently not yet been formulated by the writers proposing this interpretation.

The second of the above statements, that generalization is due to *a failure to associate* certain stimulus characteristics with the conditioned reaction, if taken literally, cannot possibly account for the *presence* of empirical generalization, but only for its *absence*.[8] Suppose, for example, that an *S* in our experiment who has been "trained" to lift his finger to the lighting of the central light fails somehow to associate finger-lifting with whatever "local signs" or distinctive cues accompany the position of that stimulus on the retina. Does it then follow that other local signs or cues arising from peripheral stimuli must now elicit finger-lifting without further training? We cannot agree that it does. Instead of observing marked generalization, we should observe *no* generalization. To appeal to the *absence* of associative connections between position cues from the central light and finger-lifting as an explanation of finger-lifting when it does occur to peripheral stimuli, would be somewhat surprising, to say the least.

But it is possible that the second statement has been misinterpreted. Perhaps its intended meaning was that the generalization of a *positive* response in a discrimination situation is due to a failure to form an association between an *avoiding or inhibiting* response and certain stimulus properties. Thus it might be contended that a subject gives a *positive* response to a nonconditioned stimulus because a competing *negative* reaction is not associated with, and hence not elicited by, the stimulus. Granting that a strong inhibitory tendency can prevent the occurrence of positive reactions, it *does not follow that a lack of inhibition can ever lead, in and of itself, to positive action;* some positive strength must also be postulated. Thus Hull (7) assumes that this conceptually necessary positive strength results from the generalization of habit strength. In the present experiment, the negative (inhibitory) tendency was presumably established by the instructions, which prohibited responses to all peripheral lights. But in spite of these instructions, the subjects exhibited a *gradient* in frequency of response to the prohibited lights. If the false reactions are to be attributed to the lack of associations between certain stimulus characteristics and the inhibitory response tendency, it then becomes necessary to explain why there was more failure-

[8] Oddly enough, Razran (10), in attempting to support the notion that much of conditioned response generalization is due to a failure of association, cites the results of six Pavlovian experiments in which not generalization, but a *lack* of generalization, was the most frequent result.

of-association for the lights near the center than for those farther removed. This would seem to require the assumption of a *gradient of failure-to-associate*, an explanatory device involving at least as many assumptions as the hypothesis it would purport to supplant.

It is also desirable to examine briefly the assertion that generalization phenomena of the sort reported here are nothing more than "discriminative thresholds [measured] under distraction" (9, p. 84). According to this point of view, the false reactions elicited by the peripheral lights in this study should not be attributed to associative connections between the response and the peripheral lights, but rather to the fact that the subjects were hurried and that hence their "limen for differential reaction under such circumstances [was] enormously increased" (9, p. 84). The steps involved in this line of reasoning are apparently (*a*) that generalization studies and psychophysical studies are fundamentally alike, since both yield discriminative thresholds, though of different sizes; (*b*) that no generalized associative strength is present in the case of psychophysical threshold determinations; and, therefore, (*c*) that no generalized habit strength is present in generalization-type studies of thresholds.

With regard to these steps, there is no question that conventional generalization studies yield data from which differential limens can be computed. Nor is there any doubt that limens computed from responses in such experiments are increased markedly over what they would have been had other procedures (e.g., classical psychophysical methods) been employed. Furthermore, a large difference limen is not unique to generalization studies; it is well known that marked variations in liminal values are obtained even within the framework of classical psychophysics as the result of variations in the subjects' attitudes, sets, or motivations, in their degree of familiarity with the stimulus dimension, in the amount of training they have had, in the temporal spacing of the stimuli, etc. We would concur, therefore, with the notion that generalization studies and psychophysical studies are basically the same. But we would reject the notion that either this fundamental similarity or the fact that generalization experiments yield large difference limens constitutes adequate grounds for dismissing as useless the assumption that positive associative strengths develop to nonconditioned stimuli during conditioning.

If, in the present experiments, the subjects had been instructed to take plenty of time in responding, as is usual in measuring threholds, it is relatively certain that no false responses (no empirical generalization) would have been observed. But this cannot be interpreted as evidence against the utility of postulating that under such conditions positive excitatory tendencies to react to the peripheral lights were present. Failure to observe an effect with a given method of measurement constitutes inadequate grounds for concluding that the effect is not present. In many instances the presence of a particular phenomenon can only be demonstrated by the use of a method of measurement that involves a delicate balance between the phenomenon being investigated and other competing processes. In the present studies, the inhibitory tendency due to the instructions not to respond to any of the peripheral lights might obscure the presence of weaker positive generalized tendencies under some condi-

tions of measurement. By urging the subjects to react quickly, it was apparently possible to introduce a general reaction set or tendency which offset the inhibitory tendency and permitted the generalized tendency to manifest itself as an empirical gradient among the equally prohibited stimuli.

In the case of the classical psychophysical methods, interest is often centered on the problem of determining the *ultimate limit of differential reactivity* attainable by a subject. In order to achieve this goal, empirical generalization, if present, must be reduced to an absolute minimum. This is accomplished by the use of highly trained subjects (where "highly trained" is surprisingly often defined in terms of a subject's ability *not* to generalize extensively, i.e. to yield a small *DL*) and by the use of "training" procedures that parallel in some respects the *differential reinforcement* procedures of the conditioning laboratory. Empirical generalization is minimized by these procedures but it is never entirely eliminated, as shown, for example, by the bidirectional curves of "equals" judgments obtained with the Method of Constant Stimuli. Now it is common practice to *theorize about* this residual variability by appealing to such factors as fluctuating receptor thresholds, shifts of attention, and inaccuracies of perception (9, p. 84). Nevertheless, it is entirely possible that some instances of poor differential reactivity, especially those due to a lack of training or to a lack of familiarity with the stimuli, might be more adequately "explained" by an appeal to the concept of habit generalization. In other words, it appears just as reasonable to assume that habit generalization is present in both generalization and psychophysical studies, and accounts in part for deviant responses, as it does to contend that it is present in neither. Hull's postulate of generalized habit strength as an explanation of broad thresholds, whether these are measured under psychophysical or generalization conditions, is certainly a more precisely defined, and hence a more scientifically useful concept, than the vague, historically antiquated notions of "shifts of attention" and "inaccuracies of perception."

Considerations such as the foregoing suggest that perhaps the relation of psychophysical studies to generalization experiments might be clarified if they were treated as the two extremes of a series of experiments arranged according to the degree to which their procedures permit the appearance of empirical generalization. At the generalization end of the series, where differential reinforcement is usually omitted or an effort is made to overcome its effects, empirical generalization is very marked, with large difference limens the rule. But at the psychophysical end, where differential reinforcement or its functional equivalent is provided by the nature of the instructions and by the use of sophisticated adult subjects, generalization is minimal and differential reactivity is exceedingly precise. The empirical generalization occurring at all points in the series might reasonably be attributed to both generalized associative strength and fluctuations in receptor-neural thresholds. The choice of one of these interpretations over the other (or perhaps the rejection of both for a third) would depend ultimately upon the relative predictive and subsumptive powers of the concepts and upon the breadth and number of meaningful relations into which they enter. Where degree of generalization can be shown to be a function of factors known or suspected

to produce fluctuations in neural or receptor sensitivity, it may be unnecessary to invoke the concept of habit generalization. But where the generalization is of an order of magnitude beyond that usually ascribed to molecular neural changes, and where it varies with more molar factors such as number of reinforcements and level of drive, the assumption of generalized habit strength may well prove more fruitful.

## Summary

Two closely parallel experiments have been carried out to determine whether a phenomenon similar to the generalization of conditioned responses to tactile stimuli would be obtained with voluntary responses to visual stimuli whose spatial positions were varied.

The stimuli consisted of seven lamps mounted in a horizontal row at eye level. Subjects were instructed to react as quickly as possible to the lighting of the middle lamp, but not to the lighting of other lamps, by lifting their fingers from a reaction key. Following an initial series of training trials in reacting to the central lamp alone, tests for the presence of generalized responses to the lighting of the peripheral lamps were introduced.

In both experiments, plots of response frequency against the spatial position of the stimulus yielded relatively smooth, symmetrical generalization gradients on each side of the central point. The response latency data appeared to show no consistent trends, though they were too irregular and unstable to permit of detailed interpretation.

The results were discussed in the light of current criticisms of generalization, an attempt being made to clarify certain terminological difficulties and to explicate, in part, the relationship of psychophysical threshold studies to generalization experiments.

(Manuscript received
December 28, 1949)

## References

1. Bass, M. J., and Hull, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, 17, 47–65.
2. Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, 33, 209–226.
3. Gibson, Eleanor J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, 24, 237–253.
4. Guthrie, E. R. Conditioning as a principle of learning. *Psychol. Rev.*, 1930, 37, 412–428.
5. Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, 17, 125–148.
6. Hovland, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, 51, 279–291.
7. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
8. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, 46, 9–30.
9. Lashley, K. S., and Wade, Marjorie. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, 53, 72–87.
10. Razran, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, 46, 337–365.
11. Razran, G. Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, 39, 642–652.
12. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430–444.

# THE INFLUENCE OF TWO VARIABLES UPON THE ESTABLISHMENT OF A SECONDARY REINFORCER FOR OPERANT RESPONSES [1]

BY PHILIP J. BERSH

*Columbia University*

This paper reports two parametric studies dealing with the acquisition by an $S^N$ of power to reinforce operants.[2] Perhaps the most widely accepted statement of a secondary reinforcement principle is that given by Hull (14, 15). As noted elsewhere (23), the statement is of restricted value without precise definitions of such terms as "close," "consistently," and "reinforcing state of affairs." Moreover, despite the central role of secondary reinforcement in reinforcement theory, few basic data concerning this phenomenon, its scope and limitations, are yet available. Recently appearing studies have begun to supply the required information (2, 3, 4, 11, 19, 21). For the most part, however, the variables and functional relationships which govern the acquisition and exercise of secondary reinforcing properties remain to be explored.

The current lack of empirical data is accompanied by a corresponding absence of explicit theory concerning secondary reinforcement. Hull (14) considers it probable that "a stimulus gradually acquires its powers of secondary reinforcement as it acquires the power of evoking the reaction conditioned to it." Hull does not specify the kind of stimulus control involved, since he recognizes only one category of behavior and one type of stimulus control. For him, the control acquirable by a stimulus over a conditioned response is elicitative in character whether the response is one like pupillary contraction or one like bar pressing. Some behavior scientists, on the other hand, follow the lead of Skinner (24) in dividing behavior into operant ("instrumental") and respondent ("classical") categories. This fundamental dichotomy leads not only to two conditioning paradigms, but to two types of stimulus control differing in history of acquisition and in their properties. Elicitative control is exercised only over respondents. Control over conditioned operants is discriminative: the stimulus "sets the occasion," or acts as a cue, for the response.

From this latter viewpoint, one could ask whether, in establishing an $S^N$ as an $S^r$ for operant behavior, the key process is acquisition by the $S^N$ of elicitative control over respondents or of discriminative control over operants. The first alternative may be

[1] A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Pure Science, Columbia University. The writer wishes to express his profound gratitude to Professor W. N. Schoenfeld, under whose constant guidance the research was performed. Special thanks are also due to Dr. J. M. Notterman for major aid in apparatus design and construction, as well as for other aid and encouragement throughout the course of these studies.

[2] For the sake of convenience, a few symbols are used throughout this paper. The following definitions are, therefore, presented:

$S^N$: A stimulus which before training has no positive reinforcing power.
$S^R$: A primary reinforcing stimulus.
$S^r$: A secondary reinforcing stimulus.
$S^D$: The positive stimulus in a discrimination procedure.
$S^\Delta$: The absence of the positive stimulus, or the negative stimulus in a discrimination procedure.

considered equivalent to a Pavlovian conditioning hypothesis of secondary reinforcement. It implies that variables will affect concomitantly the acquisition by an $S^N$ of control over respondents and of the power to reinforce operants. It does not necessarily imply, however, that the functions relating such variables to the two processes will be identical. The second alternative may be considered equivalent to an operant discrimination hypothesis of secondary reinforcement. It suggests that the process which underlies the acquisition by an $S^N$ of operant reinforcing power is its acquisition of discriminative control over operants. According to such a hypothesis, variables will affect concomitantly (though perhaps not identically) the formation of an operant discrimination and the extent to which the positive stimulus becomes an $S^r$ for operant behavior.

In the case of operants, there is evidence that an $S^N$ becomes a more powerful $S^r$ after use as positive stimulus in a discrimination procedure than after its participation in an equivalent amount of nondiscrimination training (21). Also pointing to a link between the discriminative and secondary reinforcing properties of a stimulus are the studies of Notterman (19) and Dinsmoor (2).

It is probable that a stimulus may become an $S^r$ for operants even in the absence of specific discrimination training. In the Saltzman study (21), a stimulus correlated with reinforcement on every trial, as well as one present during alternate reinforced and unreinforced trials, did acquire reinforcing properties. However, such results cannot be regarded as hostile to an operant discrimination hypothesis. Even where little or no opportunity is afforded for unreinforced responding in the absence of the potential $S^r$, the latter will gain some discriminative control over an operant because of the existence of a generalization gradient. That is, stimulus conditions present when the operant is reinforced will have greater control over that operant than will other stimulus conditions (1, 24, 27).

It must be admitted that the case for an operant discrimination, as contrasted with a Pavlovian conditioning, hypothesis is not as clear-cut as the foregoing discussion might suggest. An operant conditioning or discrimination procedure always incorporates operations presumably sufficient to give the positive stimulus elicitative control over respondents. This follows from the fact that it is paired with a reinforcing agent which elicits such respondents. Accordingly, one cannot be certain that the subsequent $S^r$ power of the stimulus is due solely or even primarily to its acquisition of discriminative control over operants. For example, in Dinsmoor's experiment (2), an $S^r$ was found to suffer an equal loss in its power to strengthen operants whether used in a discriminative (i.e., stimulus precedes the operant) or a reinforcing (i.e., stimulus follows the operant) capacity. This could be ascribed to the fact that a stimulus will tend to lose control over conditioned respondents to the same extent whenever it is not followed by primary reinforcement. The position of such a stimulus relative to a measured operant would be of no consequence in this respect.

In the final analysis, of course, whether a distinction need be made between the operant discrimination and Pavlovian conditioning hypotheses of secondary reinforcement, depends on the discovery of variables which influence acquisition of respondent-control by a stimulus in a

manner different from its acquisition of operant-control.

## Experiment I

### *The Effect of $S^N$ Duration Prior to $S^R$ Presentation*

The temporal relation of an $S^N$ to an $S^R$ is presumably one determiner of the subsequent effectiveness of the $S^N$ as a secondary reinforcer (14, 15). It now appears, however, that temporal proximity alone of an $S^N$ to reinforcing stimuli is insufficient to make the former a reinforcing agent, at least for operant responses (23). The present study deals with but one aspect of the temporal problem. The duration of a continuous $S^N$ prior to $S^R$ presentation is varied, and the effect upon acquisition by the $S^N$ of power to condition a new operant is measured. That an $S^r$ can strengthen new (i.e., previously untrained) responses has been demonstrated many times (e.g., 10, 21, 24). The principal specification of the time variable was made in terms of $S^R$ presentation as a reference event. It is, of course, necessary to establish some event in the reflex chain leading to, or involved in, reinforcement as a reference point from which to measure the time separation of a potential $S^r$, and in terms of which to identify the time relation as forward or backward. In a conditioning study employing a "free" operant, the only precise specification of time relations possible involves $S^R$ presentation as the reference point, since this is the last event in the reflex chain which is subject to experimental control. However, the possible importance of such subsequent events as seizing the food agent, onset of gustatory stimulation, and start of chewing led to the use of a procedure which would permit at least rough specification in terms of these events without changing the character of the independent variable. If $S^R$ presentation had marked the termination of the continuous $S^N$, such later events would have occurred in the absence of the latter stimulus. Instead the $S^N$ was permitted to continue for 2 sec. beyond $S^R$ presentation, a period which proved sufficient to encompass the chain at least through onset of chewing.

A second temporal variable, that involving the time separation of a *discrete* $S^N$ from $S^R$ presentation, has been investigated by Jenkins (16). Both a Pavlovian conditioning and an operant discrimination hypothesis of secondary reinforcement would apparently anticipate somewhat different consequences for the discrete and continuous cases. The indication is that the power of a stimulus to evoke a conditioned respondent decreases more sharply with CS-US interval (beyond the optimal range of interval values) for the discrete (trace) than for the continuous (delay) situation (13, 14). According to a discrimination hypothesis, a similar advantage should appear for the continuous case. In the discrete situation, the response reinforced by $S^R$ presentation always occurs in the absence of the $S^N$, which is later to be tested as a possible $S^r$; in the continuous case, especially as set up in the present experiment, the reinforced response occurs during the continuation of the potential $S^r$. For very short intervals, of course, the two cases become quite similar.

### *Procedure*

*Subjects.*—Eighty-three male albino rats, Wistar stock, 90–110 days old at the start of the experiment, were used. Of these, 23 were replacements for animals which exhibited considerable bar-biting and bar-jumping behavior during the test sessions. Actual observation of such behavior was the sole criterion for replacing an animal. Unless otherwise indicated, all data

refer only to non-biting and non-jumping animals. Shortly after their arrival at the laboratory, the rats were placed in individual living-experimental cages similar to those described in (8).

*Apparatus.*—Six sets of apparatus were used to run six animals simultaneously. Each set consisted of a response lever, two recording devices, and an arrangement for controlling the presentation and duration of a light (the $S^N$). The levers were slightly modified versions of those described in (8). Thick bars (¼ in. diameter) and false cage ceilings were introduced as the result of the difficulty with bar biting and bar jumping. Bar-pressing levels, as indicated both by operant level determinations and test sessions, were not affected by these changes. Response records were made by electrically operated counters and by four-channel, constant speed, wax tape recorders (8).

The $S^N$ in each setup was the diffuse light from a coated 6 w., 120 v. bulb. When in use, this was positioned over the food-tray end of the living-experimental cage, the end through which the bar was inserted during operant level and test sessions. With the light on, cage illumination in the food-tray area was approximately 2.7 foot-candles. Illumination during light-off periods was about .09 foot-candles. Frick (7) has demonstrated that, at this general level of illumination, a 30:1 ratio constitutes a readily discriminable pair of stimulus conditions.

The light's duration was controlled by an electronic timer, assembled from a kit manufactured by Lafayette-Concord Radio Corp., Stock No. 32-N-24518, 1948 catalogue. This permitted continuous variation of stimulus duration from .75 sec. to about 80 sec. The timing error at any duration was about 1 per cent. Once the timer had been set, light-duration was fixed. Thus, when bar depression produced the light, additional presses while the light was on could not lengthen the light period.

Temperature ranged from 74° to 80° F. during operant level and from 74° to 78° F. during test sessions.

*Operant level.*—Three days were used to begin the establishment of a hunger rhythm, consisting of 23¾ hours of deprivation and a 45-min. feeding period. On the fourth day of rhythm the response lever was made available to each rat for the last 45 min. of the twenty-third hour of deprivation. Each response produced a 1-sec. light, later to be used as an $S^N$ for pairing with the $S^R$ and then tested for secondary reinforcing properties. No experimental reinforcement of any kind accompanied bar pressing. The same procedure was used on each of the next five days. The number of responses made by each rat on days 5 and 6 were summed, and the animals were divided into six groups of ten each, equated for mean and standard deviation of these sums (22). After matching, the range for the means was from 34.8 to 35.3 responses and, for the standard deviation, from 18.1 to 19.3. The groups were assigned to experimental treatments at random.

*Training* (Six consecutive days).—Training began on the second day after the last operant-level session. On day 1, ten pellets (0.1 gm. each) were dropped one at a time and at intervals of about a minute into the food tray to train the animals to investigate the tray at the sound of a falling pellet. Then followed 20 paired presentations of a light ($S^N$) and a pellet ($S^R$). On each of five succeeding days the animals were given 28 paired presentations of light and pellet. At no time during training was the bar present. The independent variable was the duration of the light prior to pellet delivery. For the six groups the durations represented were 0, 0.5, 1, 2, 4, and 10 sec. respectively. For all groups the light was continuous from its onset until 2 sec. had elapsed after pellet delivery. Thus, the total duration of the light was in each case 2 sec. greater than the value listed above. This feature served to insure the occurrence, in the light, of such events as seizing the pellet, onset of gustatory stimulation, and onset of chewing.

Paired presentations of light and pellet were spaced at intervals ranging from 20 to about 50 sec. The spacing was varied in a non-systematic fashion to prevent the development of a rhythm. To keep the time elapsing between pellet presentation and pellet seizing as brief as possible, the light was turned on only when the animals were in the immediate vicinity of the food tray. As a precaution against fortuitous direct or inductive reinforcement of the components of bar pressing, care was taken not to present the light (and its correlated pellet) when the animal was reaching through or above the H-shaped slot through which the bar was to be inserted during test sessions.

*Test* (Three consecutive days).—This phase began on the day following the last training session. The conditions duplicated those described under the operant-level phase. Pressing responses produced a 1-sec. light, but were otherwise not reinforced. As noted earlier, light duration was fixed, so that presses which occurred within 1 sec. after light onset did not prolong the light.

## *Results and Discussion*

Table I presents the median number of bar-pressing responses made by the several groups during each of the three test sessions. Medians are used

TABLE 1

MEDIAN NUMBER OF RESPONSES MADE BY EXPERIMENTAL GROUPS DURING EACH OF THREE 45-MIN. TEST SESSIONS

| Test Session | $S^N$–$S^R$ Interval (sec.) | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 0.5 | 1 | 2 | 4 | 10 |
| 1 | 32.0 | 34.0 | 38.0 | 41.0 | 38.5 | 27.5 |
| 2 | 28.0 | 28.5 | 38.0 | 27.5 | 30.5 | 17.0 |
| 3 | 29.0 | 25.5 | 24.0 | 23.0 | 34.5 | 20.0 |

throughout the data presentation because they appear to be more representative of the sample distributions than are the means. Even with biters and jumpers eliminated, most of the samples contained a few animals whose level of responding was well out of line with that of other group members. The data for days 1 and 2 show a similar tendency to pass through a maximum in the range from 1–4 sec. On the third day the trend appears to be quite different. By this session, however, the light is no longer differentially effective as a secondary reinforcer for the several groups.

A non-parametric test (5) was used because of sample size and the lack of available information about the distributions for parent populations. Additional considerations were the presence of extreme cases in some of the sample distributions and the existence of variance heterogeneity. Nor did there seem to be any simple transformation that would normalize the sample distributions, and, at the same time, equalize their variances. Unfortunately, the test applied to the data fails to take correlation terms into account, and its use here involves an error in the direction of conservatism because of the existence of inter-group correlations (through the matches for operant level). Non-parametric tests exist which are directly applicable to correlated samples (e.g., 9, 26), but they all presuppose a specific pairing procedure which, owing to practical difficulties, could not be completely fulfilled in assigning animals to the experimental groups.

When applied to the data for the first day, the test revealed that the distribution for the 10-sec. group differed significantly (.05 level) from the distributions for the 1-, 2-, and 4-sec. groups. None of the remaining inter-group comparisons yielded values beyond the .05 level. On the second day, the distribution for the 10-sec. group was found to differ significantly at the .01 level from the distributions for the 0-, 1-, and 4-sec. groups, and at the .05 level from the distribution for the 0.5-sec. group. Other inter-group comparisons did not yield significant results. On day 3, none of the inter-group comparisons was significant.

On the basis of these results, the response totals for each animal on days 1 and 2 were combined, and medians were computed for the two-day totals. These two-day medians have been plotted in Fig. 1. The dotted curve is for the medians based on all animals, while the points through which the solid curve is drawn do not include the data for biters and jumpers.

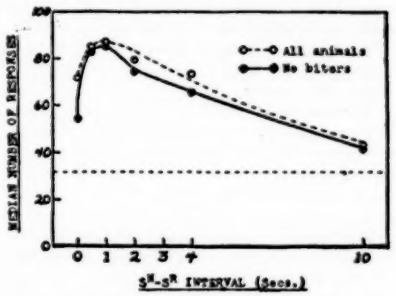


FIG. 1. Median number of bar-pressing responses made during two 45-min. test sessions combined. The horizontal dotted line is an estimate of the asymptote for the solid curve.

The horizontal line on the graph represents the estimated value of the asymptote for the solid curve. This estimate is the median number of responses during the last 1½ hr. of operant level activity (on days 5 and 6 combined of the operant-level phase).

Inter-group comparisons showed that the 10-sec. group was significantly different from the 1- and 4-sec. groups at the .01 level, and from the 0-, 0.5-, and 2-sec. groups at the .05 level. As a final check on the possibility that elimination of biters and jumpers somehow biased the data, group comparisons including all animals were also made. The result was that the 10-sec. group then differed significantly from all other groups at the .01 level.

In both sets of comparisons, none of the other inter-group differences proved significant at the .05 level. In view of the small size of the samples and the need to use a relatively insensitive test, the failure to find significant differences in the 0–4 sec. range is not surprising. It must be concluded that, while the shape and regularity of the solid curve in Fig. 1 suggest the existence of a maximum between 0 and 10 sec., this has not been definitely established by the statistical tests. Both the shape of the function and the precise location of a peak, if any exists, must be determined by replicative experiments.

Table II presents the median numbers of responses made by the experimental groups during the first 10 min. of Test Session I. While group differences are not significant, the trend follows that for the curves in Fig. 1. This suggests the existence of group differences in the reinforcing power of the test stimulus from the outset of the test phase. Accordingly, the trend of Fig. 1 is not due solely to any differential rapidity with which the test stimulus may lose its acquired reinforcing power.

TABLE II

MEDIAN NUMBER OF BAR-PRESSING RESPONSES MADE BY THE EXPERIMENTAL GROUPS DURING THE FIRST 10 MIN. OF TEST SESSION 1

| $S^N$–$S^R$ Interval (sec.) | 10 Min. |
|---|---|
| 0 | 19.0 |
| 0.5 | 25.0 |
| 1 | 25.0 |
| 2 | 24.5 |
| 4 | 22.5 |
| 10 | 12.5 |

Rho's were obtained for each group between the number of responses on days 5 and 6 combined of the operant-level phase (i.e., the matching criterion) and number of responses made on the first two test days combined. These rho's ranged from −.04 to +.66, five of the six being positive. According to the probability distribution constructed by Olds (**20**) for the sums of the squared rank differences, one of the correlations is significant at the .05 level and another almost at the .02 level. A method proposed by Fisher (**6**) for combining independent tests of significance was applied, and the aggregate was found to be significant at the .05 level. The size of the correlation for the entire set of data could not be estimated because no satisfactory procedure is available for combining rho's. Further evidence for the existence of a positive correlation between the variables will be presented in connection with Experiment II. It is this correlation which makes it desirable to match groups for operant level.

It is of interest to note some parallels between the present data and those found in studies of stimulus-response asynchronism (**14**). A summary of delay and trace conditioning studies (**12**) reveals that the optimal

CS-US interval lies in the range from 0.3 to 1.5 sec., with a majority of the reported studies placing the maximum for the function near 0.5 sec. This range is close to that which is tentatively indicated by the present data as holding for the secondary reinforcement of operants. (Since the $S^N$ in the present experiment is continuous, experiments on delay conditioning [e.g., 17] are, of course, more pertinent than those which deal with trace conditioning.) Accordingly, this study appears to provide evidence favorable to a Pavlovian conditioning hypothesis of secondary reinforcement.

The present data also relate to an attempt by Spence (25) to derive empirical gradients for delay of reinforcement. In setting up what he considers an adequate theoretical schema, he has made the following assumptions, which lead to a prediction of what the empirical curve in this experiment ought to have been:

1. Stimulus patterns which precede the pattern present at the time of primary reinforcement generalize with the latter according to a monotonic decreasing function of their time separation from it. In our procedure the interval between light onset and $S^R$ presentation was varied systematically. Light onset should, therefore, generalize with the stimulus complex present at the time of reinforcement according to the function assumed by Spence.

2. The pattern present at the time of primary reinforcement acquires $S^r$ properties directly. Earlier patterns acquire such properties according to the assumed generalization gradient. It may be inferred that, in the present experiment, light onset acquired $S^r$ power to an extent determined by such a gradient.

3. Stimuli and stimulus changes which accompany the occurrence of a response solely and directly mediate its reinforcement. In our procedure the test of secondary reinforcement involved production of light onset by the response measured.

The preceding assumptions lead to the inference that the light should have revealed $S^r$ power in the monotonic decreasing fashion indicated. This inference does not appear to be borne out by the experimental findings.

## Experiment II

### *The Effect of Number of Pairings with a Primary Reinforcing Agent*

Both a Pavlovian conditioning and an operant discrimination hypothesis of secondary reinforcement would predict that the power of an $S^N$ to reinforce operant behavior will increase as a function of its repeated correlation with an $S^R$. From the viewpoint of the latter approach, the required feature is not the mere pairing of $S^N$ and $S^R$ (a procedure sufficient in terms of the Pavlovian hypothesis), but the occurrence of operants following the onset of the $S^N$ which are then reinforced as the result of $S^R$ presentation. Unless such responses are reinforced either in the presence of the $S^N$ or following its onset, the stimulus cannot gain the discriminative control of operants demanded by the hypothesis. The fact is, of course, that the reinforcement of operants and respondents proceeds concurrently in any conditioning experiment. Such an experiment is labelled "operant" or "respondent" conditioning solely on the basis of the response being measured. As indicated in the general introduction, an operant conditioning procedure necessarily incorporates an arrangement sufficient to produce respondent conditioning. On the other side, it is difficult to conceive of a respondent conditioning experiment which would exclude all possibility of operant reinforcement. More than one observer

has remarked upon the strengthening of operants during a respondent conditioning study, even when the animals concerned have been subjected to a considerable degree of restraint (13, 28).

While the repeated pairing of $S^N$ and $S^R$ may tend to give the former some discriminative control over operants, the further development of such control requires the use of a differential reinforcement procedure (18, 24). It appears that the rapidity and completeness of $S^\Delta$ extinction hold the key to the progress of the discrimination (18, 19, 24). In the conditioning of a restricted operant (e.g., locomotion in the runway), it is possible experimentally to minimize the amount of $S^\Delta$ extinction. On the other hand, in a situation involving "free" operants (i.e., responses over which the experimenter has little or no control, like tray approach in the Skinner box), differential reinforcement is always present or possible. When $S^D$ is present, the operants occur and are reinforced by $S^R$; when $S^D$ is absent, they can also occur, but are not reinforced. Both in this experiment and in Experiment I there was ample opportunity for the differential reinforcement, with respect to presence or absence of the $S^N$, of tray approach, and nosing (23). It has already been shown that a procedure which incorporates a discrimination procedure is more effective for secondary reinforcement than one which does not (19, 21).

Since the completion of this experiment, there has appeared a study by Hall (11) which involved, among other variables, the number of $S^N$–$S^R$ pairings. The present experiment, however, covers a wider range of values and, through its use of an unrestricted operant, does not impose a ceiling upon variation of the response measure.

### *Procedure*

*Subjects.*—The subjects were 60 male albino rats, Wistar stock, 90–110 days old at the start of experimentation. One of these was discarded for biting during the extinction phase.

*Apparatus.*—The apparatus was that described in connection with Experiment I. Thick bars to reduce biting and false cage ceilings to eliminate jumping, were used throughout. The temperature ranged from 82° to 90° F. during the extinction phase and from 84° to 88° during the test phase.

*Operant level.*—The details were the same as in Experiment I. On the basis of the sums of responses for the last two days, six groups equated for mean and standard deviation were set up. After equation, the range for the means was from 38.5 to 38.8 and, for the standard deviation, from 23.1 to 24.8. Because one of its members was eliminated for biting, the control group mean increased to 41.2. The groups were assigned at random to the experimental treatments.

*Training* (Six consecutive days).—On day 1, ten pellets were dropped by hand into the food tray at intervals of about 1 min. The lever was then introduced, and pressing responses were regularly reinforced until each animal had received one-fifth of the total number of reinforcements allotted to members of his group. Members of the experimental groups received totals of 10, 20, 40, 80, and 120 reinforcements respectively. Control animals were allotted a total of 120 reinforcements. In the case of the experimental animals, each pressing response was accompanied by the onset of a 3-sec. light, pellet delivery occurring after the light had been on for 1 sec. For the controls, pellet delivery also followed bar depression by 1 sec., but was not accompanied by the light. The 1-sec. interval between light onset and pellet delivery was selected as probably optimal on the basis of the results of Experiment I. On each of the four succeeding days, bar-pressing responses were regularly reinforced until one-fifth of the total number of reinforcements for each animal had been delivered.

*Extinction* (Six consecutive days).—On each extinction day, the bar was made available for one-half hour. Responses were neither reinforced nor followed by the light. This phase was designed to reduce the strength of bar pressing to the same level for all groups. Otherwise, the effect of varying the number of light-pellet pairings upon the acquisition of $S^r$ power by the light would have been confounded with the differential effect upon bar-pressing strength produced by varying the number of primary reinforcements.

*Test* (Three consecutive days).—The lever was introduced for 45 min. per day, and responses were accompanied by a 1-sec. light, but were otherwise not reinforced.

### *Results and Discussion*

In Table III are presented the median numbers of bar-pressing responses made by the six groups during each of the three test sessions. Values for extinction day 6 are also included. Applications of a non-parametric test (5) to the data for the last extinction day failed to reveal significant group differences. For the first session, the 20- and 120-pairing groups were found to differ significantly at the .05 level, while for the second session the 120 group was significantly different (.05 level) from both the 0 and the 20 groups. There were no other significant differences.

Figure 2 constitutes a plot of the medians for response totals on the first two test days combined. The results for the third day are not included, because no trend is apparent for that day. For the two-day totals, the 120 group differed significantly (.05 level) from the 0- and 20-pairing groups. Other group differences were not significant.

Spearman rho's were computed for the six groups separately between

TABLE III

MEDIAN NUMBER OF RESPONSES MADE BY EXPERIMENTAL GROUPS DURING THE SIXTH EXTINCTION SESSION (30 MIN.) AND DURING EACH OF THREE 45-MIN. TEST SESSIONS

| Session | $S^N$-$S^R$ Pairings | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 10 | 20 | 40 | 80 | 120 |
| Extinction 6 | 17.0 | 16.5 | 19.5 | 13.5 | 23.0 | 20.5 |
| Test 1 | 20.0 | 18.5 | 18.5 | 27.0 | 27.0 | 30.0 |
| 2 | 13.0 | 17.5 | 18.0 | 22.0 | 26.5 | 33.5 |
| 3 | 19.0 | 16.5 | 17.5 | 21.5 | 20.0 | 20.0 |

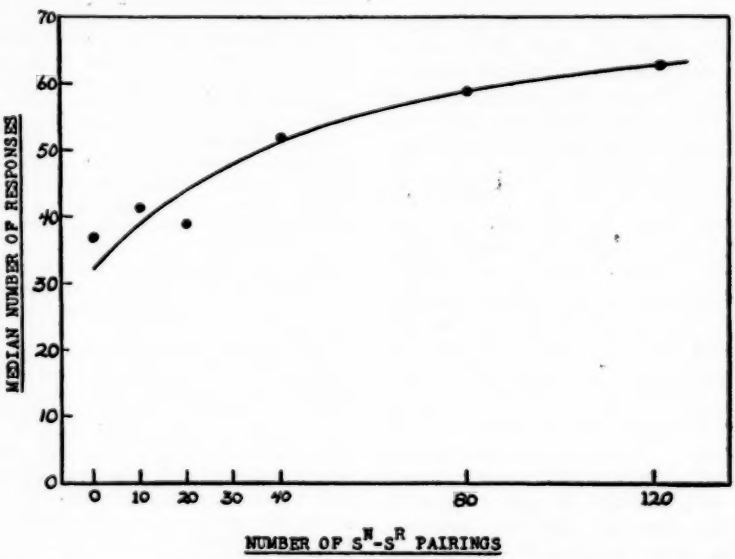


FIG. 2. Median number of bar-pressing responses made during two 45-min. test sessions combined.

operant-level criterion values of individual rats and number of responses made by the rats on the first two test days combined. The range was from .16 to .91. Combining the probabilities for the six tests of significance (6) resulted in .01 level significance for the aggregate. In the course of the two experiments reported here, 12 sample correlations between operant level and test session responding were obtained, of which 11 were positive. Evaluated by a binominal expansion, the probability of obtaining such a result on the null hypothesis is less than .004.

The procedure used here was substituted for that of Experiment I partly because it was considered more likely to reveal the influence of a small number of pairings upon the acquisition of $S^r$ power by an $S^N$. The test situation in Experiment I involved the conditioning of a previously untrained operant, whereas in Experiment II the test situation requires only the reconditioning of a response. An examination of Fig. 2, however, reveals little or no superiority for the 10- and 20-pairing groups over the control group. Furthermore, even the 80-pairing group failed to differ significantly from the control group.

Two considerations seem to bear upon the magnitude of the effect obtained:

1. During training, the spacing of successive $S^N$–$S^R$ pairings was under the control of the animal. Whenever it depressed the bar, light onset and pellet delivery followed. As the number of primary reinforcements accumulated, one bar press tended to follow another closely. The consequent massing of $S^N$–$S^R$ pairings might be considered unfavorable to the development of secondary reinforcement from the viewpoint of either a Pavlovian conditioning or an operant discrimination hypothesis. According to the latter, massing is unfavorable because there is less opportunity for $S^\Delta$ extinction of tray approach and nosing to occur.

2. The stimulus complex with respect to which tray approach and nosing were trained discriminatively consisted, not of light alone, but of light plus proprioceptive stimuli from bar pressing. This follows from the fact that the former responses were reinforced only in the presence of the light, which, in turn, depended for its onset upon bar depression. During the post-training dark-extinction of bar pressing, the proprioceptive stimuli were present but were never accompanied by reinforcement. Thus, they tended to lose whatever generalized control over tray approach and nosing they had gained as part of the $S^D$ complex during training. This might be expected to exert a large inductive extinction effect upon the discriminative properties of the entire complex. An analogous interpretation in terms of Pavlovian conditioning is clearly possible.

For the values in this experiment which overlap the range investigated by Hall, the function appears to be as flat and as linear as the one he obtained. In addition, it has the overall negative acceleration of his curve when extrapolated by him to a zero number of $S^N$–$S^R$ pairings. For the range from 0–20, the slope of the curve reported here is also quite low. In view of the foregoing discussion, it is possible that this is an artifact of the procedure employed.

## Summary

1. In Experiment I the duration of a neutral stimulus prior to the pres-

entation of a primary reinforcing agent was systematically varied, and the effect of such variation upon the acquisition by the neutral stimulus of power to reinforce a previously untrained operant was measured. Six groups of ten rats each were matched with respect to pre-training bar-pressing activity. Each group was then given 160 paired presentations of a continuous light and a pellet. For the six groups the durations of the light before pellet delivery were 0, 0.5, 1, 2, 4, and 10 sec. respectively. Subsequently, the lever (absent during training) was again made available, and each bar press produced the training light for 1 sec.

2. The obtained curve, based upon bar-pressing totals for 1.5 test hr., suggests a maximum in the range between 0 and 2 sec. The curve did not appear to have attained its asymptote at 10 sec.

3. The relation of these results to findings for S–R asynchronism and to Spence's derivation of delay of reinforcement gradients is pointed out.

4. Experiment II was a parametric study to determine the influence of the number of pairings of a neutral stimulus with primary reinforcement upon acquisition by the stimulus of power to reinforce operants. Five groups of ten rats each and a control group of nine rats were matched for pre-training bar-pressing activity. The experimental groups were given respectively 10, 20, 40, 80, and 120 regular reinforcements for bar pressing. Each response produced a 3-sec. light, reinforcement occurring after the light had been present for 1 sec. The control group was given 120 regular reinforcements. Each reinforcement occurred 1 sec. after bar depression but was not accompanied by the light. The strength of bar pressing was then reduced to the same level for all groups by a period of extinction in the dark. When this had been achieved, each bar depression (for all animals) was permitted to produce the training light for 1 sec.

5. A negatively accelerated curve of low initial slope was obtained for number of bar-pressing responses made during 1.5 test hr. The 120 pairing group differed significantly (.05 level) from the 0- and 20-pairing groups. Other group differences were not significant.

6. In the two experiments combined, 12 Spearman rho's were obtained between number of responses made during the last 1.5 hr. of pre-training activity (i.e., the matching criterion) and during the first 1.5 test hr. Of these, 11 were positive. Evaluated by a binomial expansion, these results are significant at the .004 level.

(Manuscript received
January 19, 1950)

## References

1. Cook, D. An experiment on discrimination training without selective reinforcement. Unpublished Master's thesis, Columbia Univ. Library, 1949.
2. Dinsmoor, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, **40**, 458–472.
3. Estes, W. K. Generalization of secondary reinforcement from the primary drive. *J. comp. physiol. Psychol.*, 1949, **42**, 286–295.
4. Estes, W. K. A study of motivating conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1949, **39**, 306–310.
5. Festinger, L. The significance of differences between means without reference to the frequency distribution function. *Psychometrika*, 1945, **11**, 97–105.
6. Fisher, R. A. *Statistical methods for research workers.* (10th Ed.) Edinburgh: Oliver and Boyd, 1946.
7. Frick, F. C. An analysis of an operant discrimination. *J. Psychol.*, 1948, **26**, 93–123.

8. Frick, F. C., Schoenfeld, W. N., and Keller, F. S. Apparatus designed for introductory psychology at Columbia College. *Amer. J. Psychol.*, 1948, 61, 409–414.
9. Friedman, M. The use of ranks to avoid the assumption of normality. *J. Amer. statist. Ass.*, 1937, 32, 675–701.
10. Grindley, G. C. Experiments on influence of amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929–30, 20, 173–180.
11. Hall, J. F. A study of the strength of a secondary reinforcing stimulus as a joint function of the number of primary reinforcements, strength of drive, and characteristics of the stimulus. Unpublished Ph.D. dissertation, Ohio State Univ., 1949.
12. Harrison, J. M. Hull's derivation of stimulus asynchronism: A correction. *Psychol. Rev.*, 1949, 56, 252–260.
13. Hilgard, E. R., and Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
14. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
15. Hull, C. L. Behavior postulates and corollaries. Mimeographed memorandum on file at Institute of Human Relations, Yale Univ., November, 1949.
16. Jenkins, W. O. A temporal gradient of derived reinforcement. *Notes*, Conference on Experimental Analysis of Behavior, Note No. 19, November, 1949, issued from Indiana Univ. (Mimeo.)
17. Kappauf, W. E., and Schlosberg, H. Conditioned responses in the white rat. III. Conditioning as a function of the length of the period of delay. *J. genet. Psychol.*, 1937, 50, 27–45.
18. Keller, F. S., and Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
19. Notterman, J. M. The interrelationships among aperiodic reinforcement, discrimination learning, and secondary reinforcement. Unpublished Ph.D. dissertation, Columbia Univ., 1950.
20. Olds, E. G. Distributions of sums of squares of rank differences for small numbers of individuals. *Ann. math. Statist.*, 1938, 9, 133–148.
21. Saltzman, I. J. Maze learning in the absence of primary reinforcement: A study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, 42, 161–173.
22. Schoenfeld, W. N., Antonitis, J. J., and Bersh, P. J. Unconditioned response rate of the white rat in a bar pressing apparatus. *J. comp. physiol. Psychol.*, 1950, 43, 41–48.
23. Schoenfeld, W. N., Antonitis, J. J., and Bersh, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1950, 40, 40–45.
24. Skinner, B. F. *The behavior of organisms.* New York: D. Appleton Century, 1938.
25. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, 54, 1–8.
26. Wilcoxon, F. Individual comparisons by ranking methods. *Biometric Bull.*, 1945, 1, 80–83.
27. Winnick, W. A. Response vacillation in conflict situations. Unpublished Ph.D. dissertation, Columbia Univ., 1949.
28. Zener, K. The significance of behavior accompanying salivary secretion for theories of the conditioned response. *Amer. J. Psychol.*, 1937, 50, 384–403.

# PROBLEM SOLVING AS A FUNCTION OF THE SITUATION[1]

BY MELVIN R. MARKS

*Tulane University*

Problem solving has two aspects—the problem itself (P), and the *behavior* of the problem solver (B). The writer believes that, for fruitful investigation, P must meet certain criteria, the reasons for which will be clarified in a proposed theory of problem solving. These criteria are: (a) plausibility (for interest and motivation); (b) complexity and difficulty (for elicitation of variable B); (c) structure (for quantification, and the recasting of P in different contexts); (d) solubility. It is suggested that if P be quantified in terms of its *elements* (the implicit or explicit "givens" of P), and if B be quantified in terms of behavior relating to these elements, a functional relationship between P and B may be demonstrated.

The core of the theory is awareness by *S* of the elements. To the extent that he is aware, *S* will be likely to solve P. A convenient index of subjective awareness is vocalization of an element. If *S* is "forced" to be analytical, awareness is facilitated, and conversely if analysis is disrupted. The "forcing" of analysis may be accomplished by a Socratic questioning technique; analysis may be presented to *S* "ready-made" in the form of a list of P elements. P may be presented to *S* spontaneously and loosely formulated in an attempt to simulate a "real-life" situation, or it may be made highly specific in simulation of a textbook exercise. The disruptive effect of a novel situation is offered as a reason for the failure of transfer of training in problem solving—lack of analysis making for lessened awareness of elements, which in turn makes solution improbable. The hypotheses tested experimentally in this study will be clearer if they are stated after a description of the procedures.

[1] A condensation of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Tulane University. The writer wishes to express his gratitude to Professor Cecil W. Mann and Harry M. Johnson for their advice and help.

## Procedure

All *S*s participating were first taught a method for computing square roots with a calculating machine and an auxiliary table of factors. The method consists of: (a) adding into the machine the number for which the root is sought; (b) adding to this a similar number obtained from the table; (c) dividing the sum by a divisor factor obtained from the table. The quotient is the desired root. *E* had falsified tabular divisor entries in some instances so that when these were used in computation an incorrect root of approximately half the magnitude of the true root was obtained. Every *S* was allowed to complete four practice exercises to familiarize himself with the method.

All *S*s served in either a personal (simulated real-life) or an impersonal (simulated textbook) situation. In the *personal* situation, *S*, after completing his training, was given four test-exercises, the second and fourth of which involved false table entries. P arose when *S* was made aware of his "errors" and was told to ascertain the source of error. In the *impersonal* situation, *S* was given a typed narrative concerning one John Jones who had been trained in the method, had solved practice exercises, and had failed to get correct roots for the second and fourth of four test exercises. P arose here when *S* was asked what *he* would do if he were in John Jones' place and had to correct the errors. It should be noted that P in either case has but four elements, and that they are discrete. The elements are: (1) the *operator* (*S* himself or John Jones) who might be repeating a manipulative error; (2) the calculating *machine* which might be making mechanical mistakes on some numbers; (3) the *method*, which might lack generality; (4) the divisor *table*, which might contain typographical or computational errors. It is

possible to eliminate tentatively (1) after demonstrable accuracy with some problems; (2) may be eliminated absolutely by showing identity of results when pencil and paper operations which duplicate machine operations are performed; (3) may be eliminated absolutely by algebraic demonstration; an examination of (4), however, would reveal inconsistencies which would lead to solution; e.g., the divisor factor appropriate to 26.3 began with the sequence, "191 . . . ," whereas divisors relevant to the companion numbers 26.1 and 26.5 began with the sequence, "101. . . ." (*E* had in this case changed the "0" to a "9" in the second digit of the divisor.)

In addition to the personal or impersonal situation, the variables, training, list, and aid, were introduced with particular groups. Training was dichotomized as lectured or nonlectured. *Lectured Ss*, about three weeks prior to experimental sessions, were given as part of a regular meeting of a general psychology class, training in "correct" problem solving. The "correct" way involved the analysis of *any* P into its elements, the determination of the requirements for solution, the formulation and testing of hypotheses, etc. There were included examples of the technique, *viz.*, "Should I go home for Christmas Holidays?" "Should I get married before I graduate?" "How may one develop a formula for the area of any triangle?" *Nonlectured Ss* received no special training.

The list variable was dichotomized as list or nonlist. For the *list Ss*, after the first 10 min. of B, there was furnished a typed list of the four elements of the problem, together with the statement that the source of the error could be attributed to one of these four. *Nonlist Ss* proceeded without such assistance. The aid variable was dichotomized as aid or nonaid. With the *aid Ss* throughout B, *E* asked at intervals, "I want you to ask yourself, 'What are the elements of my problem? What are the only ways in which a mistake could occur?' " If *S* strayed from the analytical attitude engendered by the questions he was brought back to it by repetition. If he vocalized some elements he was encouraged to test them and search for more. *E* never named elements or indicated them by gesture.

Two dependent variables were considered—vocalizations and solutions. *S* was given credit for a *vocalization* if he spoke the name of an element aloud or pointed to it, together with some statement which indicated his belief that the particular element might be a source of error. From a pilot study, a manual of typical phrases used by *Ss* was constructed, together with scoring criteria. *S*'s "best," i.e., most unequivocal, vocalization of any given element was the one considered for scoring purposes; only one vocalization per element was credited for any given *S*, so the vocalization score range was 0 to 4. An *S* was given credit for a *solution* if he showed that the second digit of the divisor was in error, supplied the correct digit, and repeated the computations with the correct divisor to secure the true root. A total of 25 min. of B was allowed for each *S*. This time was split into a 10- and a 15-min. period with a 2-min. interpolated break in which *S*'s vital statistics were recorded. Vocalizations were scored from verbatim protocols transcribed from a Peirce Wire Recorder. Vocalizations were not scored unless they occurred in the first period.[3] Additional apparatus included a typed exposition of the method for obtaining square roots; a modified, mimeographed copy of a square root divisor table published by the Marchant Calculating Machine Co.; typed numbers in sets of four, each number consisting of two digits to the left and one digit to the right of the decimal, all sets being different; a typed exposition of the predicament of John Jones for use in the impersonal situation.

Within the meaning of the definitions, and with the phrase, "All other factors assumed to vary at random," implied in every instance, the experimental hopotheses stated in the form of *predictions* are as given below. It should be noted that Hypotheses I–III are concerned with vocalizations, whereas IV–VIII refer to solutions. Note also that, for II, V, and VII, the theory predicts that the null hypothesis will remain tenable—it is for this reason that these hypotheses are given formally.

I. Impersonal situations will evoke more vocalizations than personal.

II. Lectured and nonlectured groups will not differ in vocalizations.

III. Aid groups will evoke more vocalizations than nonaid.

IV. Impersonal situations will evoke more solutions than personal.

V. Personal and impersonal situations will not differ in solutions *if* a list of elements is furnished to *S*.

VI. List groups will evoke more solutions than nonlist.

VII. Lectured and nonlectured groups will not differ in solutions.

VIII. Aid groups will evoke more solutions than nonaid.

In all, 82 *Ss* participated in the experimental tests of these hypotheses. These were dis-

[3] To tap adequately *S*'s awareness of elements, after any 15-sec. period of silence, *E* encouraged him to vocalize by saying, "Tell me what you are thinking about. Say it no matter how irrelevant you may believe it to be."

tributed in three groups with *N*s of 18, 42, and 22. From the description of the groups given below, it will be seen that for certain of the hypotheses, *S*s drawn from more than one of the three groups could be used for statistical comparisons.[3]

Group I ($N = 18$) was used to compare the personal and impersonal situations without list, training, or aid. *S*s were drawn from the 87 Arts and Sciences Freshmen who scored between the fortieth and sixtieth percentiles on the Ohio State Psychological Examination. After 47 veterans or athletes living in dormitories were excluded, letters of invitation were issued by the college dean, and the first 18 who responded were used. Two groups of nine each were formed by matching the OSPE scores for mean and standard deviation, and one of the groups was assigned to the personal situation on the basis of a coin toss.

Group II ($N = 42$) was used to compare simultaneously situations and training, with list given in every instance, but no aid. Of the 42, 22 *S*s drawn from a general psychology quiz section had been lectured on problem solving about three weeks prior to the experiment proper. Two groups of 11 each were obtained by simultaneously matching for mean and standard deviation, the "concept-formation," "vocabulary," and "total" scores yielded by the Shipley-Hartford Retreat Scale, which had been administered to the entire class. Allocation of the subgroups to situations was by coin toss. The remaining 20 *S*s in Group II were drawn from an experimental psychology class; two groups of ten each were matched as above on the Shipley-Hartford Scale. Allocation to situations was by coin toss. None of these 20 had been lectured. To recapitulate, Group II consisted of four subgroups: lectured-personal ($N = 11$); lectured-impersonal ($N = 11$); nonlectured-personal ($N = 10$); nonlectured-impersonal ($N = 10$).

Group III ($N = 22$) was used to compare aid and nonaid in a personal situation when no training or list was given. Two groups of 11 each were drawn from an experimental psychology class on the basis of Shipley-Hartford scores as in Group II.

[3] For example, the comparison of personal and impersonal situations with respect to vocalizations involves all *S*s from Group I, the nonlectured personal and impersonal *S*s from Group II, and the nonaid personal *S*s from Group III; a comparison of list and nonlist with respect to solutions involves all *S*s from Group II (since lecture has no demonstrable effect) for list, and all *S*s of Groups III and I for nonlist, with the exception of the aided *S*s of Group III.

## Results

Original data are summarized in Table I. For the statistical comparisons the vocalizations were dichotomized: those *S*s scoring 3 or 4 vocalizations classed as *upper*; those scoring 1 or 2 as *lower*. Results are summarized by hypotheses.

Hypothesis I is tenable. In personal situations 8 of 41 *S*s were upper, while 18 of 30 impersonal were upper. Corrected $\chi^2 = 10.55$, $d.f. = 1$, $P < 0.001$.

Hypothesis II (the null) is tenable. Of nonlectured *S*s, 6 of 20 were upper, while of lectured *S*s, 10 of 22 were upper. Corrected $\chi^2 = 1.12$, $d.f. = 1$, $P > 0.20$.

Hypothesis III is tenable. Only 5 of 30 nonaid *S*s were upper, while 10 of 11 aid *S*s were so classed. Corrected $\chi^2 = 16.06$, $d.f. = 1$, $P < 0.001$.

Hypothesis IV is tenable. Of personal situation *S*s only 8 of 41 solved P, while 16 of 30 impersonal *S*s solved P. Corrected $\chi^2 = 7.41$, $d.f. = 1$, $P < 0.01$.

Hypothesis V (the null) is tenable. When all *S*s were given a list, in the personal situation 4 of 21 *S*s solved P, and in the impersonal situation 11 of 21 solved P. Corrected $\chi^2 = 3.73$, $d.f. = 1$, $P > 0.05$ (cf. Discussion).

Hypothesis VI fails. Although it was predicted that list *S*s would show more solutions than nonlist *S*s, 9 of 29 nonlist *S*s solved P, and 15 of 42 list *S*s solved P. Corrected $\chi^2 = 0.03$, $d.f. = 1$, $P > 0.80$. An interpretation is suggested in the discussion.

Hypothesis VII (the null) is tenable. Of nonlectured *S*s, 5 of 20 solved P, and of lectured *S*s, 10 of 22 solved P. Corrected $\chi^2 = 1.12$, $d.f. = 1$, $P > 0.20$.

Hypothesis VIII is tenable. Of nonaid *S*s only 5 of 30 solved P, while

TABLE I

FREQUENCY OF VOCALIZATIONS AND SOLUTIONS BY SUBGROUPS

| | Group I Nonlist-Nontrained | | Group II List | | | | Group III Personal-Nonlist-Nontrained | |
|---|---|---|---|---|---|---|---|---|
| | Personal | Impersonal | Lectured | | Nonlectured | | Aid | Nonaid |
| | | | Personal | Impersonal | Personal | Impersonal | | |
| *N* | 9 | 9 | 11 | 11 | 10 | 10 | 11 | 11 |
| Vocalizations | 19 | 28 | 24 | 30 | 18 | 25 | 35 | 18 |
| Solutions | 1 | 5 | 3 | 7 | 1 | 4 | 11 | 3 |

all 11 aid *Ss* solved P. Corrected $\chi^2 = 20.12$, $d.f. = 1$, $P < 0.001$.

It was predicted also that there would be a positive relationship between vocalization and solution. If the data for vocalization be classed as upper and lower as described above, and that for solutions as solved and nonsolved, the assumptions for tetrachoric correlation are met. The strength of the relationship is indicated by $r_{tet} = 0.83 \pm 0.23$. Other tetrachoric *r*'s are given in Table II.

TABLE II

TETRACHORIC CORRELATION COEFFICIENTS FOR RELATIONSHIP BETWEEN VARIOUS DEPENDENT AND INDEPENDENT VARIABLES

| Variables | $r_{tet}$ | *N* |
|---|---|---|
| Vocalizations—impersonal | .65 | 82 |
| Solutions—impersonal | .58 | 82 |
| Solutions—list (aid *Ss* excluded) | .09 | 71 |
| Lecture—vocalizations | .45 | 42 |
| Lecture—solutions | .45 | 42 |
| Aid—vocalizations | .94 | 52 |
| Aid—solutions | .82 | 52 |

All are consistent with the theory. Coefficients for relationships between the dependent variables and the external matching criteria are given in Table III. They are uniformly low.

There are some miscellaneous results which deserve mention. (1) The frequencies of *Ss* scored as vocalizing 1, 2, 3, or 4 of the elements were distributed almost perfectly as a binomial expansion of $p = .5$ with $n = 3$. No *S* failed to vocalize at least one element. (2) No combination of elements, whether taken two at a time, three at a time, etc., occurred significantly more frequently than any other, i.e., for a given *S* vocalizing of a given element did not lead to a regular pattern of other elements. This is surprising, since at first thought it would seem that attention paid to elements other than the critical one would tend to make solution less likely. Results indicate that this is not the case. (3) Personal and impersonal situations showed similar patterning in the relative frequency of vocalization of the four elements, in each case the operator being vocalized significantly less frequently and the machine significantly more frequently than chance would predict. (4) The experimental groups did not differ with respect to ease of vocalization

TABLE III

RELATIONSHIPS BETWEEN DEPENDENT VARIABLES AND EXTERNAL CRITERIA

| Variables | Coefficient | *N* |
|---|---|---|
| Solutions vs. Shipley-Hartford | $r_{bis} = .44$ | 42 |
| Vocalizations vs. Shipley-Hartford | $r = .30$ | 42 |
| Vocalizations vs. Ohio State P.E. | $r = .46$ | 20 |

as measured by number of times *E* had to ask *S* to vocalize during the first 10 min. of B. (5) The experimental groups did not differ in amount of stereotyped behavior when such was defined as number of non-useful repetitions of a calculating machine process, i.e., when *S* continued to repeat a calculation even though it yielded an incorrect root.

## Discussion

The superiority of the impersonal situation over the personal is more marked than the data would indicate. It will be remembered that these were intended to simulate textbook exercises and real-life problems respectively. Neither a true textbook nor a true real-life situation is reproducible in an investigation of this kind. The former has real-life aspects in that *S* performs an unfamiliar task in the presence of an unfamiliar person. The latter has textbook aspects in that *S* knows he is participating in an artificial situation. In this sense the personal and impersonal situations are "closer" together than their intended counterparts, and the obtained differences are more striking.

The demonstrated superiority of the Socratic questioning in evoking both vocalizations and solutions raises the question of whether the analytical attitude engendered by the questioning can be transferred. In view of the failure of the training in problem solving to transfer in this study, the writer doubts whether such transfer can ever be achieved. It may be thought that the training lacked intensiveness and extensiveness, i.e., a one-hour lecture might not show demonstrable effects three weeks later. But routine academic training in "thinking," though persisting through four years, is highly tenuous. Can *it* be expected to transfer to a real-life situation, say twenty years later? Additional light may be shed on this point if the data are considered in another way. In addition to the all-or-none criterion used for solutions, a so-called "fine-scoring" was employed as well. This permitted partial scores for solutions, e.g., an *S* who attributed the error to the table but could not localize it further received more credit than one who had no idea as to the scource of the error after the 25 min. of B. These results were not reported because, with one exception, they paralleled the all-or-none results. However, the exception occurred with the training variable. For the "fine-scoring," solutions for the lectured group exceeded those for the non-lectured group at the 5 per cent level. The transfer effect seems anomalous in that it appears or does not appear as a function of the scoring. However, if the all-or-none criterion had been less rigorous, e.g., if attributing the error to the table was sufficient for a score of "solved," then that system too would indicate the superiority of the lectured group. It is suggested that something which might be called "solution activity" is transferred—*S*s who are trained tend to reach some solution, albeit that solution may not be complete. The extension to real-life problems is obvious.

Although the theory predicted that the list of elements would substitute for analytical awareness, the experimental results were disappointing. There was no difference in solutions between list and nonlist *S*s, and the impersonal-list *S*s were superior to the personal-list *S*s at close to the 5 per cent level of confidence. (The obtained $\chi^2 = 3.73$, whereas for the 5 per cent level, 1 *d.f.*, required $\chi^2 = 3.84$.) The following interpretation

is offered. *S* becomes aware of the elements not through his own efforts, but artificially. The inadequacy of the list may be analogous to the failure of highly directive advice. If P may be considered as composite, then *S* must be aware of the elements *and* he must apply such awareness. Thus awareness and application are severally necessary and together sufficient for solution. When the list is given, there may be intellectual awareness without attentional awareness. It is suggested that the disruptive effect of the personal situation is such that *S* does not attend to the elements of the list.

The relationship between vocalization and solution ($r_{tet} = 0.83$) was demonstrated. There is reason to believe that an even higher coefficient would have been attained had not an experimental artifact spuriously depressed the correlation. *Ss* who considered the factor table early in the 10-min. period during which vocalizations were scored, tended to restrict their attention to it throughout the balance of the period; usually, it led to solution. In fact, if the table were considered at any time except last with reference to the other elements, and if solution followed, then the other elements were precluded from awareness and vocalization. Thus, the total number of elements vocalized was reduced, and this reduction was particularly great with *Ss* who solved P. The resulting restriction of range of vocalization scores, together with the tendency for solvers to have fewer vocalizations (in the relative sense described above), lowered the correlation.

## Summary and Conclusions

It is proposed that there is a positive relationship between *S*'s awareness of the givens of a problem (herein called the *elements*) and his ability to solve that problem. Awareness depends on analysis, and loosely formulated (*personal*) problem situations tend to disrupt analytical capacity, whereas concisely formulated (*impersonal*) problem situations do not. *Ss* were taught a method of extracting square roots with the use of a calculating machine and an auxiliary factor table. *S*'s problem in a personal situation arose when he made errors in test exercises because *E* had falsified the table, and *S* was asked to find the source of the error. The problem in the impersonal situation arose when *S* was asked to solve John Jones' problem—John being a character in a written description of a personal situation. For either *S*, the elements of the problem were the *S*, the machine, the table, and the method. All 82 *Ss* used were allowed 25 min. of work consisting of a 10-min. period, a 2-min. break, and a 15-min. period. The index of awareness was vocalization of elements, and these were scored from the verbatim protocols of the 10-min. period. Criterion for solution was finding and correcting the misprint.

Forty-two *Ss* were divided into a personal and an impersonal situation group. Eleven of the 21 in each group were lectured on problem solving three weeks prior to the solving sessions. All 42 received a list of the four problem elements at the start of the 15-min. period. An additional 18 *Ss* were split into personal and impersonal situation groups. None of these was given the list. An additional 22 *Ss* in a personal situation were split into an aid and a nonaid group of 11 each. The aid group was forced to analyze, through Socratic questions.

The results appeared to justify the following conclusions:

1. The impersonal situations were superior to personal in both vocalizations and solutions (1 per cent level).

2. The lectured and nonlectured groups did not differ significantly in either vocalizations or solutions, though the lectured were slightly better.

3. For the groups furnished with a list of the elements, a statistically insignificant difference in favor of the impersonal group with respect to solutions was found.

4. The aid group was markedly superior to nonaid in vocalizations and solutions (1 per cent level).

5. The relationship between vocalization and solution is indicated by $r_{tet} = 0.83 \pm 0.23$.

(Manuscript received
January 19, 1950)

PSYCHOLOGICAL MONOGRAPH

#270

1945

# ON PROBLEM SOLVING

By

KARL DUNCKER

$2.50

This popular monograph is #270 of the Psychological Monograph series. It has been reprinted so that it is again available.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N. W.

Washington 5, D. C.

# AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W.
Washington 5, D. C.

The American Psychological Association, founded in 1892 and incorporated in 1925, is the major psychological organization in the United States. With approximately 8,500 members and 2,000 affiliates, it includes most of the qualified psychologists in the country. The purpose of the APA is to advance psychology as a science, as a profession, and as a means of promoting human welfare. It attempts to further these objectives by holding annual meetings, publishing psychological journals, conducting a personnel placement service for psychologists, and working toward improved standards for psychological training and service.

In order to give recognition to the specialized interests of different psychologists, the APA includes seventeen Divisions. Any person, after becoming a member of the APA, may apply for membership in as many Divisions as he wishes. The Divisions are:

Division of General Psychology
Division on the Teaching of Psychology
Division of Experimental Psychology
Division on Evaluation and Measurement
Division on Childhood and Adolescence
Division of Personality and Social Psychology
The Society for the Psychological Study of Social Issues—a Division of the American Psychological Association
Division on Esthetics
Division of Clinical and Abnormal Psychology
Division of Consulting Psychology
Division of Industrial and Business Psychology
Division of Educational Psychology
Division of School Psychologists
Division of Counseling and Guidance
Division of Psychologists in Public Service
Division of Military Psychology
Division on Maturity and Old Age

Each division has its own officers. Each meets annually at the time and place of the APA meeting. Each has its own membership requirements, which in some cases are higher or more specialized than the requirements for election to the APA.

The annual meetings of the APA are held in September. An attempt is made to choose meeting sites in such a way that attendance will be convenient for members and their guests in different sections of the country at different times. Thus, the 1951 meeting will be held from August 31 through September 5 in Chicago, and the 1952 meeting in Washington, D. C. These meetings provide the members with an opportunity to present or hear reports of psychological research and lectures on topics of psychological interest, and to participate in formal and informal discussion groups. The research reports, lectures, and discussion groups cover varied topics of current interest within the field of psychology.

The chief governing body of the APA is the Council of Representatives. It includes representatives from each of the Divisions. A Board of Directors, composed of the four officers of the APA (President, President-elect, Recording Secretary, and Treasurer) and six Council members elected by the Council, is the administrative agent of the Council and exercises general supervision over the affairs of the Association.

The APA maintains an office in Washington which acts as a coordinating center for all APA activities. Journal subscriptions, membership applications, personnel placement requests, and general APA business procedures are handled in this office.